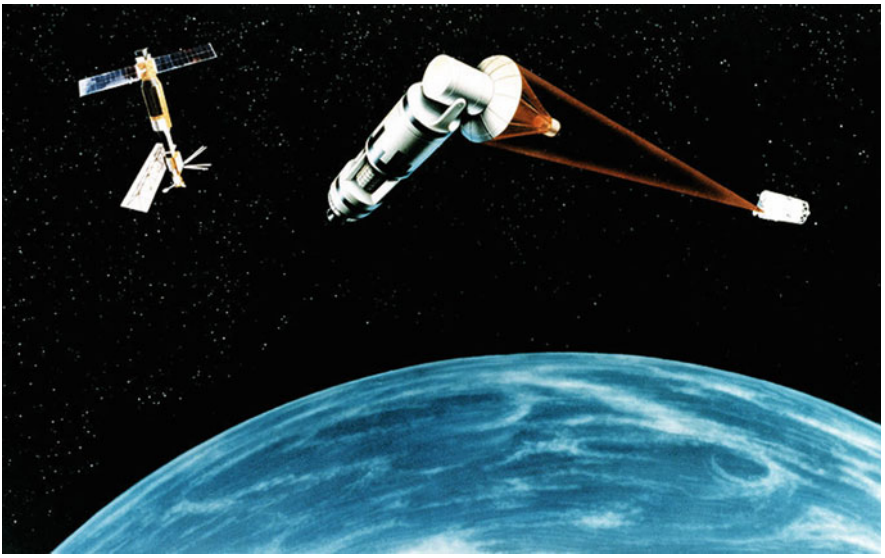
Bahman Zohuri

Directed Energy Weapons

Physics of High Energy Lasers (HEL)

Springer

Directed Energy Weapons



Bahman Zohuri

Directed Energy Weapons

Physics of High Energy Lasers (HEL)

Springer

Bahman Zohuri
Galaxy Advanced Engineering, Inc.
Albuquerque, NM, USA

ISBN 978-3-319-31288-0 ISBN 978-3-319-31289-7 (eBook)
DOI 10.1007/978-3-319-31289-7

Library of Congress Control Number: 2016946373

© Springer International Publishing Switzerland 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

This book is dedicated to my mother and father Marzieh and Akbar Zohuri.

Without their encouragement, this book would not have been written.

Preface

Directed energy weapons are nothing new to mankind; historically the origination of such weapons began centuries ago when the famous Greek mathematician, physicist, engineer, inventor, and astronomer; Archimedes of Syracuse used different mirrors to collect sunbeams and focused them on the Roman fleet in order to destroy enemy ships with fire. This is known as the Archimedes Heat Ray. Archimedes may have used mirrors acting collectively as a parabolic reflector to burn ships attacking Syracuse. The device was used to focus sunlight onto approaching ships, causing them to catch fire. Of course the myth or reality of the Archimedes Heat Ray still is questionable, but with the help of a group of students from Massachusetts Institute of Technology certain experiments were carried out with 127 one-foot (30 cm) square mirror tiles in October of 2005 that were focused on a mockup wooden ship at a range of about 100 feet (30 m). The flames broke out on a patch of the ship, but only after the sky had been cloudless and the ship had remained stationary for around 10 min. It was concluded the device was a feasible weapon under these conditions.

The battles of tomorrow will be fought with different weapons that have more lethal effects and faster delivery systems. One of mankind's greatest achievements in the twentieth century is the ability to destroy the entire human race several times over. At this time of intensive arms more money is invested in the next generation of weapons. It is in the best interest of every citizen to be aware and able to make an informed judgment on the best possible direction for the arms race. Offensive or defensive weapons are a cruel reality that nevertheless must be addressed.

The scientific work during the 1950s that led to the invention of the laser was followed closely by work in military research institutes and organizations all over the world and this opened a new door to the Archimedes Heat Ray. Lasers have found many military applications, not as new weapons, but rather as the supporting technology to enhance the performance of other weapons such as laser-guided bombs and so on. Our fascination and appreciation of modern weaponry is at an all-time high. It was not until the 1970s that the possibility of laser weapons again

captured the imagination of military planners. High-energy and other directed energy weapons finally became a reality, and the possibility of using them in the battlefields of tomorrow has been investigated vigorously ever since.

The development of laser weaponry and other directed energy weapons technology conjures up the Heat Ray of Archimedes and Flash Gordon-like images of vaporizing enemies, demolishing buildings, and burning through metal. In this book introduces such weaponry to readers of different technical backgrounds as well as to introduce a certain technical approach to such research and to help better understanding of such weapons utilizing various technical and research resources.

The next 10 years will see the emergence of high-energy lasers as an operational capability in US service. These weapons will have the unique capability to attack targets at the speed of light and are likely to impair the effectiveness of many weapon types significantly, especially ballistic weapons. Constrained by propagation physics, these weapons will not provide all-weather capabilities, and will perform best in clear sky–dry air conditions.

The book in its laser technology section talks about the interaction between high-power laser beams and matters whereas other aspects of directed energy weapons, such as particle and high-power radar beams as a weapons of tomorrow can be found in the literature provided by other authors. Laser-beam interactions with materials, treat, from a physicist's point of view, the wide variety of processes that lasers can induce in materials. Physical phenomena ranging from optics to shock waves are discussed. The approach that is taken emphasizes the fundamental ideas both from a newcomer's or research worker's point of view to provide important background for material science, mathematics, optics, and the like, or a most critical up-to-date review of the field.

A directed energy weapon (DEW) such as a high-energy laser emits energy in an aimed direction without the means of a projectile. It transfers energy to a target for a desired effect. Some such weapons are real or in development; others are at present only in science fiction.

The energy can come in various forms:

- Electromagnetic radiation (typically lasers or masers)
- Particles with mass (particle beam weapons)
- Sound (sonic weaponry)
- Fire (flamethrowers)
- High-power laser weapons

Some lethal directed energy weapons are under active research and development, but most examples appear in science fiction, nonfunctional toys, film props, or animation.

In science fiction, these weapons are sometimes known as death rays or ray-guns and are usually portrayed as projecting energy at a person or object to kill or destroy. Many modern examples of science fiction have more specific names for directed energy weapons, due to research advances.

For those readers who need to dive deep into the technologies behind such research a short course in various topics of mathematics and physics has been offered in the appendices in order for them to brush up on these topics and be able to understand different solutions and mathematical modeling that are offered for the solution, for example, of the heat diffusion equation for different boundary and initial conditions. In the case of application of lasers as weapons, the book has attempted to serve both scientists interested in the physical phenomena of laser effects and engineers interested in practical applications of laser effects in industry. Thus, several sections are devoted to reviewing and dealing with the solution of the diffusion equation utilizing the aid of integral transform techniques. Among the several different approaches to solve the boundary value problems for heat conduction; the integral transform technique offers the most straightforward and elegant solution, provided that the transforms, the inversions, and the kernels are readily available.

Some appendices at the end of the book are devoted to systematic mathematics and physics of the heat conduction solution and its boundary value problems. As a result of the transforms, the inversions, complex variables, and their examples are presented and the kernels are tabulated, and the Laplace and Fourier transforms are also introduced. The appendix on introduction to ordinary and partial differentials is also presented to help the reader understand the solution techniques used to solve the heat conduction problem for various boundary values. Appendices on optics and the electromagnetic field also help better understanding of the behavior of the physics and mathematics of these weapons.

Note: In most of the appendices of different topics either the references mentioned at the end of each appendix have been used and quoted directly or indirectly or it is up to each reader to refer to them separately for more knowledge and information. I have also decided to shift these appendices around by eliminating some of their content that I believe is no longer necessary, as well as converting some content into part of the main chapters of different subjects of Volume II here and finally keeping the rest as appendices as originally planned.

Those left as an appendix of their own for those readers needing some refresher and review on the topics that are presented by these appendices are:

Appendix A: Short Course in Taylor Series
Appendix B: Short Course in Vector Analysis
Appendix C: Short Course in Ordinary and Partial Differential Equations
Appendix D: Short Course in Complex Variables
Appendix E: Short Course in Fourier and Laplace Transforms
Appendix F: Short Course in Electromagnetics
Appendix G: Short Course in Optics
Appendix H: Short Course in Heat Conduction Equation
Appendix I: Data and Plots of Thermal Parameters of Different Materials
Appendix H: Acronyms and Definitions

In this book, I have also taken under consideration to show the solutions and present the heat conduction complex problem and those boundary values that are very much related to problems of high- power laser interaction with materials. Most cases have looked at one-dimensional heat conduction with semi-infinite slab configuration with a heat resource as part of heat conduction equations making dealing with it a more difficult and complex problem. Wherever was needed the best possible references were given for further investigations by readers interested in doing their own research beyond what is given here.

Albuquerque, NM Bahman Zohuri

Acknowledgments

I am indebted to the many people who aided and encouraged me and the people who supported me beyond expectations. Some of them are not around to see the end result of their encouragement in the production of this book, yet I hope they can see this acknowledgment. I especially want to thank Nancy Reis of Sandia National Laboratory who both put the idea in me to start the book based on some work that I was contracted to do for them. My thanks go also to Joe Rogers of NASA, retired now, one of my best friends who helped with most of the computer codes that are presented in this book to bring them to their present status from their legacy stages and to be able to transfer them from a mainframe to a personal computer platform under the Windows operating system.

Another best friend, William Kemp of the Air Force Weapons Laboratory at Albuquerque, New Mexico is really a true friend and remains as one. Finally my many thanks to Jonathan W. Plant, Senior Editor—Mechanical, Aerospace, and Nuclear and Energy Engineering of Taylor & Francis/CRC Press who made all this happen. Finally, I am indebted to many people and the individuals and organizations that granted permission to reproduce copyright materials and published figures.

I am also indebted to Dr. John T. Schriemf and his time on the phone on so many occasions and advice about the different topics related to high-power laser interaction with matte, in particular allowing me to tap into his well-known Naval Research Laboratory (i.e., NRL Report 7728) under the title of "Response of Materials to Laser Radiation a Short Course," and allowing me to copy some of his well-written report to reflect on in this volume as presented in Chapter 1.

I also would like to thank Dr. Krzysztof Nowakowski for letting me use his efforts and research work and allowing me to utilize some sections of his research work. His work and his research under the title of "Laser Beam Interaction with Materials for Microscale Applications" was most helpful to me and Chapter 7 of the book.

Above all, I offer very special thanks to my mother and father while they were alive, and my wife and children. They provided constant interest and encouragement, without which this book would not have been written.

Their patience with my many absences from home and long hours in front of the computer during preparation of the manuscript is especially appreciated.

Contents

1 Directed Energy Weapons 1
1.1 Introduction 1
1.2 PUFF74: A Material Response Computer Code 4
1.2.1 Availability of PUFF74 Computer Code 5
1.3 PUFF-TFT: A Material Response Computer Code 6
1.3.1 Availability of PUFF-TFT Computer Code 10
1.4 SANDYL: A Monte Carlo Three-Dimensional Computer Code 10
1.4.1 Availability of SANDYL Computer Code 13
1.5 ASTHMA88 (Axisymmetric Transient Heating and Material Ablation) Code 13
1.5.1 Availability of ASTHMA88 Computer Code 14
1.6 ALE3D (Arbitrary Lagrangian/Eulerian Multi-Physics 3D) Computer Code 14
1.6.1 ALE3D Program Availability 17
1.7 CTH Computer Code 17
1.7.1 Availability of CTH Computer Code 20
1.8 HYPUF, Stress Wave Response Computer Code 20
1.8.1 Availability of HYPUF, Stress Wave Response Computer Code 21
1.9 DYNA2D and DYNA3D Computer Codes Series 21
1.9.1 Availability of DYNA2D and DYNA3D Computer Codes 22
1.10 NIKE2D and NIKE3D Computer Codes Series 23
1.10.1 Availability NIKE2D and NIKE3D Computer Codes Series 24
1.11 TOPAZ2D and TOPAZ3D Computer Codes Series 24
1.11.1 Availability TOPAZ2D and TOPAZ3D Computer Codes Series 25
References 26

2 Laser Technology 27
2.1 Basic Principles 27
2.2 Overall Theme 28
2.3 A Word About Units 29
2.4 Developing Damage Criteria 29
2.5 The Energy Required for Damage 29
2.6 The Laser Beam 30
2.7 Summary 33
References 33

3 Laser Safety 35
3.1 Laser Safety 35
3.2 Laser Hazards 36
3.2.1 Laser Hazards to the Eye 36
3.2.2 Laser Hazards to the Skin 39
3.3 Safety Regulations 39
3.4 Laser Hazard Classification 40
3.5 Laser Range Safety Tool (LRST) Physics 41
References 46

4 Laser Weapons 47
4.1 Laser as a Weapon 47
4.2 Possible Targets 48
4.3 Energy Level at the Target 48
4.4 Absorption and Scattering 50
4.5 Atmospheric Structure with Altitude 52
4.6 The Major Laser Weapon Concepts 53
4.7 Small-Scale Weapons Using Lab-Type Lasers 55
4.8 High-Energy Lasers as Weapons 55
4.9 High-Energy Laser (HEL) Safety Program 56
4.9.1 Airborne Laser (YAL-1A) 58
4.9.2 Tactical High-Energy Laser for Air Defense 63
4.10 Lasers for Air Defense 64
4.10.1 Target Acquisition for Combat Operations 67
4.10.2 Overview 67
4.10.3 Description 67
4.11 Target-Background Discrimination for Surveillance 70
4.11.1 Overview 70
4.11.2 Description 71
References 77

5 Laser-Directed Energy Concepts 79
5.1 Laser Beam and Material Interactions and Its Lethality 79
5.2 Introduction to Effectiveness of Directed Energy Weapons 81
5.3 The Mathematics of Diffusion 82

5.3.1 The Diffusion Process and Basic Hypothesis of Mathematical Theory 83
5.3.2 The Differential Equation of Diffusion Equation 86
5.3.3 Boundary and Initial Conditions 89
5.3.4 Material Response 91
5.4 Effects Caused by Absorption of Laser Radiation at the Surface 133
5.4.1 Heating Without Phase Change 138
5.4.2 Heating with Phase Change 139
5.4.3 Melt-Through of a Metal Plate 141
References 142

6 High-Energy Laser Beam Weapons 145
6.1 Introduction 145
6.2 Directed Energy Weapons Engagements 149
6.2.1 Acquisition, Tracking, Pointing, and Fire Control 153
6.3 Wavelength Effects 154
6.4 The Atmospheric Propagation Problem 158
6.4.1 Laser Light Scattering and Intensity 162
6.5 Thermal Blooming Effects 168
6.5.1 Mathematical Foundation of Thermal Blooming 171
6.6 Adaptive Beaming and Imaging in Turbulent Atmosphere 172
6.6.1 Adaptive Optics 176
6.6.2 Deformable Mirror 178
6.6.3 Large Optical Systems 181
6.6.4 What Is Phase Conjugation in Optics? 181
6.7 Target Effects 191
6.7.1 Measured Characteristic of Target Both Optically and Thermally 192
6.7.2 Target Absorptance Optical Approach 193
6.7.3 Target Absorptance Thermal Approach 194
6.7.4 Mathematical Modeling of Thermal Approach 196
References 198

7 Lasers 201
7.1 Introduction 201
7.2 How Laser Works 206
7.3 Laser Light Propagation 212
7.4 Physics of Laser Absorption in Metals 215
7.4.1 Description of the Phenomena 217
7.5 The Behavior of Electromagnetic Radiation at Interface 218
7.5.1 Light Propagation in Materials 221
7.5.2 Depth of Focus 228
7.5.3 Laser Beam Quality 230
7.5.4 Spherical Aberration 231
7.5.5 Thermal Lens Effect 232

7.6 Theoretical Discussion of Laser Absorption and Reflectivity 233
7.6.1 Reflectivity of Materials at Infrared Wavelength 239
7.7 Mathematical of Laser Absorption in Metals 250
7.8 Material and Thermal Response 258
7.8.1 Boundary Conditions 260
7.9 Solutions of Governing Equation 264
7.9.1 Analytical Methods 264
7.9.2 Melting Process 331
7.9.3 Melting and Vaporization 338
7.9.4 Electron–Phonon Analytical Solution 344
7.10 Comparison of Fourier and Kinetic Theory 351
7.11 Finite Difference Methods 352
7.12 Effects of Pulsed Wave Laser Radiation 353
7.12.1 Power Levels of Pulsed Wave Laser 353
7.12.2 Material Vaporization Effects 353
7.12.3 Effects from Absorption of Radiation in the Plume 358
7.13 Effects of Continuous Wave Laser Radiation 373
References 376

8 Atmospheric Propagation of High-Energy Laser Beams 379
8.1 Introduction 379
8.2 Laser Propagation in the Atmosphere 382
8.2.1 Cloud Descriptions 384
8.2.2 Absorption and Scattering of Laser Beam by Gases and Solids 385
8.3 Laser and Thermal Blooming Effects 397
8.4 Mission Impact 405
8.5 Adaptive Optics 407
8.6 Current Initiatives 411
References 412

Appendix A: Short Course in Taylor Series 415

Appendix B: Short Course in Vector Analysis 423

Appendix C: Short Course in Ordinary and Partial Differential Equations 453

Appendix D: Short Course in Complex Variables 535

Appendix E: Short Course in Fourier and Laplace Transforms 585

Appendix F: Short Course in Electromagnetic 685

Appendix G: Short Course in Optics 751

Appendix H: Short Course in Heat Conduction 767

Appendix I: Data and Plots of Thermal Parameters of Different Materials 795

Appendix J: Acronyms and Definitions 809

Index 813

About the Author

Bahman Zohuri is currently with Galaxy Advanced Engineering, a consulting company that he started himself in 1991 when he left both the semiconductor and defense industries after many years of working as a chief scientist. After graduating from the University of Illinois in the fields of physics and applied mathematics, he joined Westinghouse Electric Corporation where he performed thermal hydraulic analysis and natural circulation for the inherent shutdown heat removal system (ISHRS) in the core of a liquid metal fast breeder reactor (LMFBR) as a secondary fully inherent shut system for secondary loop heat exchange. All these designs were used for nuclear safety and reliability engineering for a self-actuated shutdown system. He designed the mercury heat pipe and electromagnetic pumps for large pool concepts of LMFBR for the heat rejection purpose for this reactor around 1978 where he received a patent for it. He later was transferred to the defense division of Westinghouse where he was responsible for the dynamic analysis and method of launching and handling of the MX missile out of the canister. The results are applied to MX launch seal performance and muzzle blast phenomena analysis (i.e., missile vibration and hydrodynamic shock formation). He was also involved in analytical calculation and computation in the study of the nonlinear ion wave in rarefying plasma. The results are applied to the propagation of the "soliton wave" and the resulting charge collector traces, in the rarefactions characteristic of the corona of a laser-irradiated target pellet. As part of his graduate research work at Argonne National Laboratory Zohuri performed computation and programming of the multiexchange integral in surface physics and solid-state physics. He holds different patents in areas such as diffusion processes and design of a diffusion furnace while he was senior process engineer working for different semiconductor industries such as Intel, Varian, and National Semiconductor. Later on he joined Lockheed Missile and Aerospace Corporation as Senior Chief Scientist and was responsible for in R&D and the study of vulnerability, survivability, and both radiation and laser hardening of different payload components (i.e., IR sensor) for the defense support program (DSP), boost surveillance and tracking satellite (BSTS), and space surveillance and tracking satellite (SSTS) against laser or

nuclear threat. While there he also studied and performed the analysis of characteristics of laser beam and nuclear radiation interaction with materials, transient radiation effects in electronics (TREE), electromagnetic pulse (EMP), system-generated electromagnetic pulse (SGEMP), single-event upset (SEU), blast and thermomechanical, hardness assurance, maintenance, and device technology.

He spent several years consulting for his company Galaxy Advanced Engineering with Sandia National Laboratories (SNL), where he supported development of operational hazard assessments for the Air Force Safety Center (AFSC) in concert with other interested parties. The intended use of the results was their eventual inclusion in Air Force instructions (AFIs) specifically issued for directed energy weapons (DEW) operational safety. He completed the first version of a comprehensive library of detailed laser tools for the airborne laser (ABL), advanced tactical laser (ATL), tactical high-energy laser (THEL), and the mobile/tactical high-energy laser (M-THEL), among others.

He also was responsible for SDI computer programs involved in battle management C^3 and artificial intelligence (AI), and autonomous systems. He is the author of several publications and holds various patents such as "Laser Activated Radioactive Decay and Results of Thru-Bulkhead Initiation."

Chapter 1
Directed Energy Weapons

Will the United States develop laser and beam weaponry for a strong nuclear defense to replace the policy of mutually assured destruction. Has the Soviet Union violated treaties by using "yellow rain" in Afghanistan and Indochina? What future lies in store for the clean neutron bomb? What kinds of super missiles are being tested for the future? What new biological and chemical weapons has the United States been cooking up?

1.1 Introduction

The idea of using an omnipotent "death ray " on the battlefield is not a new concept. Ancient literature credits the Greek mathematician Archimedes as the first to conceive the idea of using light as a defensive weapon. Hippocrates, commander of the Greek force, applied Archimedes' concept by focusing the energy of sunlight through a series of mirrors to produce a beam that set fire to the sails of the Roman fleet under Consul Marcus Claudius Marcellus during the siege of Syracuse in 212 BC. [1]

Laser weapon projects have been shrouded by very tight security. In spite of this, it is possible to follow the general lines, at least, of the high-energy laser (HEL) weapon research field through the open literature.

Our fascination and appreciation of modern weaponry is at an all-time high. With the wonders and horrors of the Persian Gulf war and event of 9/11 fresh in our minds, the development of laser weapon technology conjures up Flash Gordon-like images of vaporizing enemies, demolishing building, and burning through metals [2].

Armed forces in many countries are already using a great number of laser devices, and the inexorable pace of progress in developing special laser weapons indicates that they will ultimately revolutionize the modern battlefield. Practically invisible when fired, silent, capable of pinpoint accuracy, and traveling at the speed

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_1

of light, laser weapons would seem to offer unparalleled advantages over conventional weapons.

This paper examines the effects of lasers—one of the first exotic directed energy weapons (DEWs) to capture public attention—and our focus in particular is on airborne laser (ABL).

While this may be true, once unleashed, this awesome power can trigger other devastating consequences as well. Anti-eye laser weapons are currently being developed that can result in the mass blinding of soldiers, pilots, and tank crews.

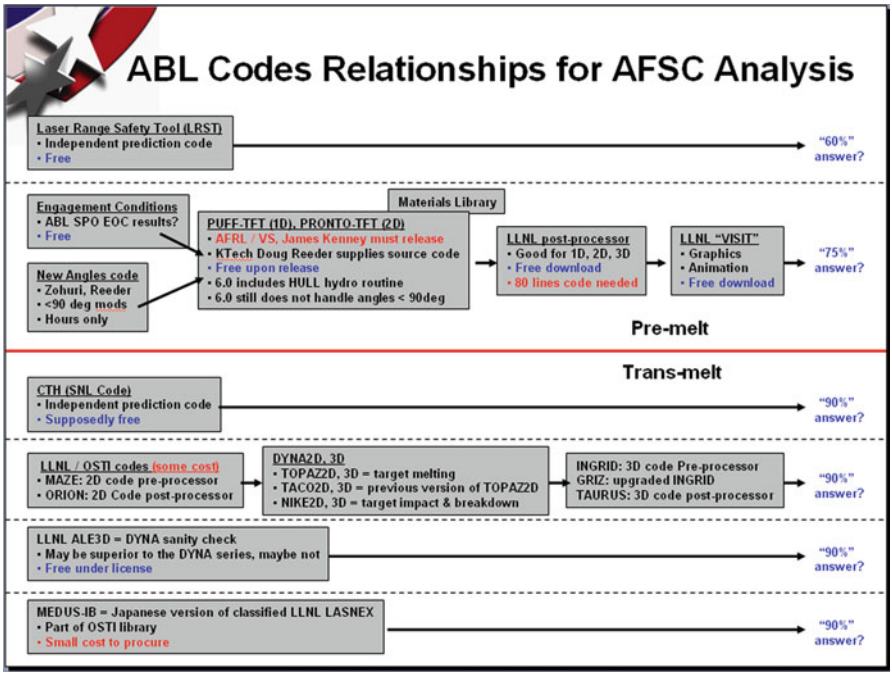
Laser experts Anderberg and Wolbarsht [2] describe the staggering medical, social, and psychological ramifications that the use of laser weapons will entail and becomes as part of our study for laser range safety tool (LRST), which is the subject of this project. Furthermore, we explore the historical development of these weapons and briefly touch the fact that how far other countries, including France, England, and Russia, have progressed in their technology. We also try to serve the purpose of an introduction to the language of directed energy weapon (DEW) for military planners and nontechnical persons who need them to understand the fundamental of what the engineers and scientists involved in their development are talking about by basically touching some physics and mathematics involved in this field. Describe all the difficulties these folks are dealing with and how to overcome these obstacles in order to produce the right tool and technologies that will induce the proper DEW for our defense of our country. A better weapon in hands of our arm force to carry on such defensive task. Any employee found to have violated these policies and procedures may be subject to disciplinary action, up to and including termination of employment.

We have put a collection of software and computer codes that are developed by our engineers and scientist around the nation within national laboratories and defense companies under one umbrella and developed the Windows/PC version of these codes as a source of repository of such capabilities. Most of these codes were developed on main- and macro-frame computing system, and under this project, we managed to migrate them to micro-frame environment for simplicity of running these codes for the purpose of further enhancement and development of LRST and DEW in ABL area. In some cases, we further enhanced the technical capability of these legacy codes in order to serve the purpose of today's technology toward development of the directed energy weapons. Table 1.1 is a good representation of these codes.

Folks who are interested to obtain these codes for the purpose of their work toward the field of DEW should contact either principal of this report or original source that is mentioned in Tables 1.1 and 1.2 in this chapter. Although we tried to gather all existing unclassified computer code together, they are still considered critical military information that requires proper paper work to obtain these computer codes from our data bank, and in some cases it might require licensing query of third-party software.

To obtain most of these codes, you are required to have some direct and related US government contract from appropriate office or agency with the government.

Table 1.1 ABL code relationships for AFSC analysis


ABL Codes Relationships for AFSC Analysis
Laser Range Safety Tool (LRST)
• Independent prediction code
• Free
"60%" answer?
Engagement Conditions
• ABL SPO EOC results?
• Free
New Angles code
• Zohuri, Reeder
• <90 deg mods
• Hours only
Materials Library
PUFF-TFT (1D), PRONTO-TFT (2D)
• AFRL / VS, James Kenney must release
• KTech Doug Reeder supplies source code
• Free upon release
• 6.0 includes HULL hydro routine
• 6.0 still does not handle angles < 90deg
LLNL post-processor
• Good for 1D, 2D, 3D
• Free download
• 80 lines code needed
LLNL "VISIT"
• Graphics
• Animation
• Free download
"75%" answer?
Pre-melt
Trans-melt
CTH (SNL Code)
• Independent prediction code
• Supposedly free
"90%" answer?
LLNL / OSTI codes (some cost)
• MAZE: 2D code pre-processor
• ORION: 2D Code post-processor
DYNA2D, 3D
• TOPAZ2D, 3D = target melting
• TACO2D, 3D = previous version of TOPAZ2D
• NIKE2D, 3D = target impact & breakdown
INGRID: 3D code Pre-processor
GRIZ: upgraded INGRID
TAURUS: 3D code post-processor
"90%" answer?
LLNL ALE3D = DYNA sanity check
• May be superior to the DYNA series, maybe not
• Free under license
"90%" answer?
MEDUS-IB = Japanese version of classified LLNL LASNEX
• Part of OSTI library
• Small cost to procure
"90%" answer?

Then the offices that are in charge of these codes will arrange for the release of these codes.

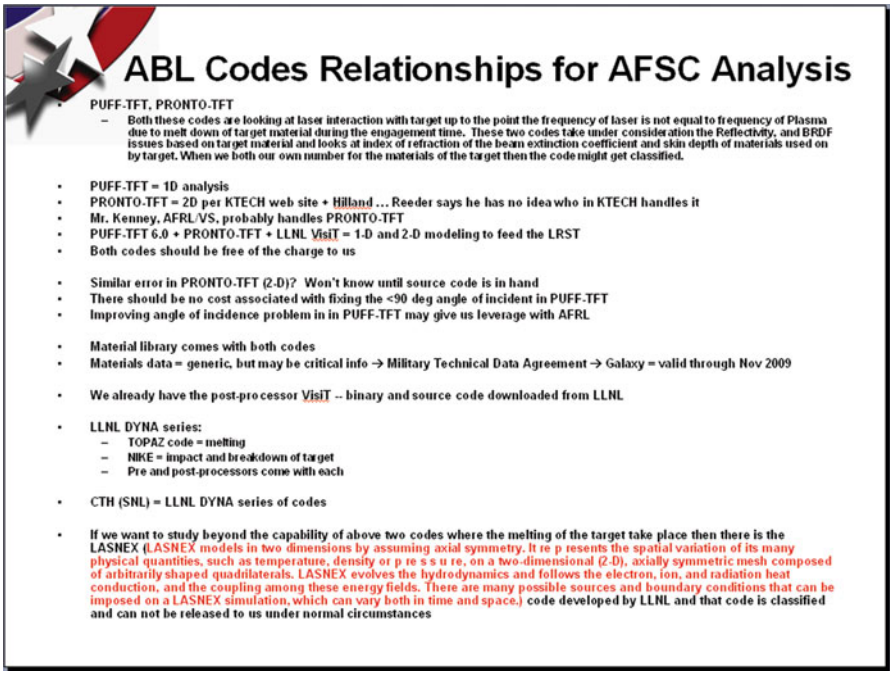
For example, most of codes that are released from Sandia National Laboratory such as CHT or codes such as DYNA3D and other related codes from Lawrence Livermore National Laboratory fall in these categories.

Although the author of this book has most of these codes, in most cases we have to follow the federal government guidelines for the release of these codes as well minus few exceptions. Please get in touch with the author for further discussion of these codes as well as consult his web site at www.gaeinc.com for availability of some of these codes through his company.

We welcome any comment and correction by expert in this field to correct and enhance our assumption on these codes or recommend better computational analysis and software tools.

In the next few sections of this chapter, we introduce unclassified computer codes that may be obtained from the source that is known to this author. Some of these codes although not classified but restricted and are available to US government agencies or their contractors.

Table 1.2 Comments on ABL code relationships for AFSC analysis



ABL Codes Relationships for AFSC Analysis

- PUFF-TFT, PRONTO-TFT
  - Both these codes are looking at laser interaction with target up to the point the frequency of laser is not equal to frequency of Plasma due to melt down of target material during the engagement time. These two codes take under consideration the Reflectivity, and BRDF issues based on target material and looks at index of refraction of the beam extinction coefficient and skin depth of materials used on by target. When we both our own number for the materials of the target then the code might get classified.

- PUFF-TFT = 1D analysis
- PRONTO-TFT = 2D per KTECH web site + Hilland … Reeder says he has no idea who in KTECH handles it
- Mr. Kenney, AFRL/VS, probably handles PRONTO-TFT
- PUFF-TFT 6.0 + PRONTO-TFT + LLNL VisiT = 1-D and 2-D modeling to feed the LRST
- Both codes should be free of the charge to us

- Similar error in PRONTO-TFT (2-D)? Won't know until source code is in hand
- There should be no cost associated with fixing the <90 deg angle of incident in PUFF-TFT
- Improving angle of incidence problem in in PUFF-TFT may give us leverage with AFRL

- Material library comes with both codes
- Materials data = generic, but may be critical info → Military Technical Data Agreement → Galaxy = valid through Nov 2009

- We already have the post-processor VisiT -- binary and source code downloaded from LLNL

- LLNL DYNA series:
  - TOPAZ code = melting
  - NIKE = impact and breakdown of target
  - Pre and post-processors come with each

- CTH (SNL) = LLNL DYNA series of codes

- If we want to study beyond the capability of above two codes where the melting of the target take place then there is the LASNEX (LASNEX models in two dimensions by assuming axial symmetry. It re p resents the spatial variation of its many physical quantities, such as temperature, density or p re s s u re, on a two-dimensional (2-D), axially symmetric mesh composed of arbitrarily shaped quadrilaterals. LASNEX evolves the hydrodynamics and follows the electron, ion, and radiation heat conduction, and the coupling among these energy fields. There are many possible sources and boundary conditions that can be imposed on a LASNEX simulation, which can vary both in time and space.) code developed by LLNL and that code is classified and can not be released to us under normal circumstances

1.2 PUFF74: A Material Response Computer Code

The **PUFF74** code is a computer code, which calculates stress wave formation and propagation by numerical integration of the conservation equations in a one-dimensional Lagrangian coordinate system. The code has been under development since 1961 and has evolved from a simple hydrodynamics code to a flexible material response code, which includes the effects of material strength, porosity, and fracture for both homogeneous and composite materials.

The code at present version (Version 4.0) is capable of handling the following physical models:

1. A framework for calculating the material response in composite materials
2. A pressure-volume-energy equation of state model for homogeneous materials or constituents of a composite material
3. A pore-compaction model for porous homogeneous materials or constituents of a composite materials
4. A one-dimensional viscoplastic model for geometric dispersion effects in composite materials

The latest model development for the **PUFF74** code has been accomplished under the CADRE program. As part of this program, studies have been made to determine the dynamic material properties, which govern the response of composite materials to rapid energy deposition.

To facilitate the input procedures for radiation deposition calculations, an automatic initial zoning model was added to the PUFF 66 code in extensive use of the code. Reference [3] presents the description of the automatic zoning model 1969 by Cooper. The guidelines used to develop this model evolved through.

The next addition to the PUFF code was the framework for introducing a free surface into the sample mesh at a location where material fracture is detected. The logic for introducing free surfaces, calculating the response of free surfaces as a function of time, and recombining fractured segments was a developed model in the original coding. Since the coding was written in a modular form, more sophisticated fracture models could be substituted with a minimum of effort.

As model development and calculations of experimental tests continued, a graphics package was added to the PUFF code to allow the user to produce online plots and externally produced plots and data storage (Calcomp or microfilm). Galaxy Advanced Engineering, Inc. (GAE) has used the Universal Graphics products known as UGL to replace the CA-DISSPLA and produce all the following graphics out and show the power of UGL for its CA-DISSPLA compatibilities. The plot package added to the code made extensive use of the general graphics data display programs developed at the Air Force Weapons Laboratory (AFWL) display program existing at ARIL at the time the plot package was added to PUFF. Modifications to the AFWL data display program and to the graphics package in the PUFF code have been made on a continuing basis to improve the efficiency of the plotting procedures.

The following is a random selection of graphics output of PUFF74/UGL combined (Fig. 1.1).

1.2.1 Availability of PUFF74 Computer Code

The Windows/PC version of this code is available from Galaxy Advanced Engineering, Inc. for purchase price. This version has been modified from its original version that used to run on VAX/VMS computer, and users need to obtain their own copy from this company. Contact the Galaxy Advanced Engineering, Inc. for its Windows/PC version. To our knowledge the code is no longer available from government agencies or its contractors. You can find more detail on how to obtain and purchase by referring to the following URL: https://www.gaeinc.com.

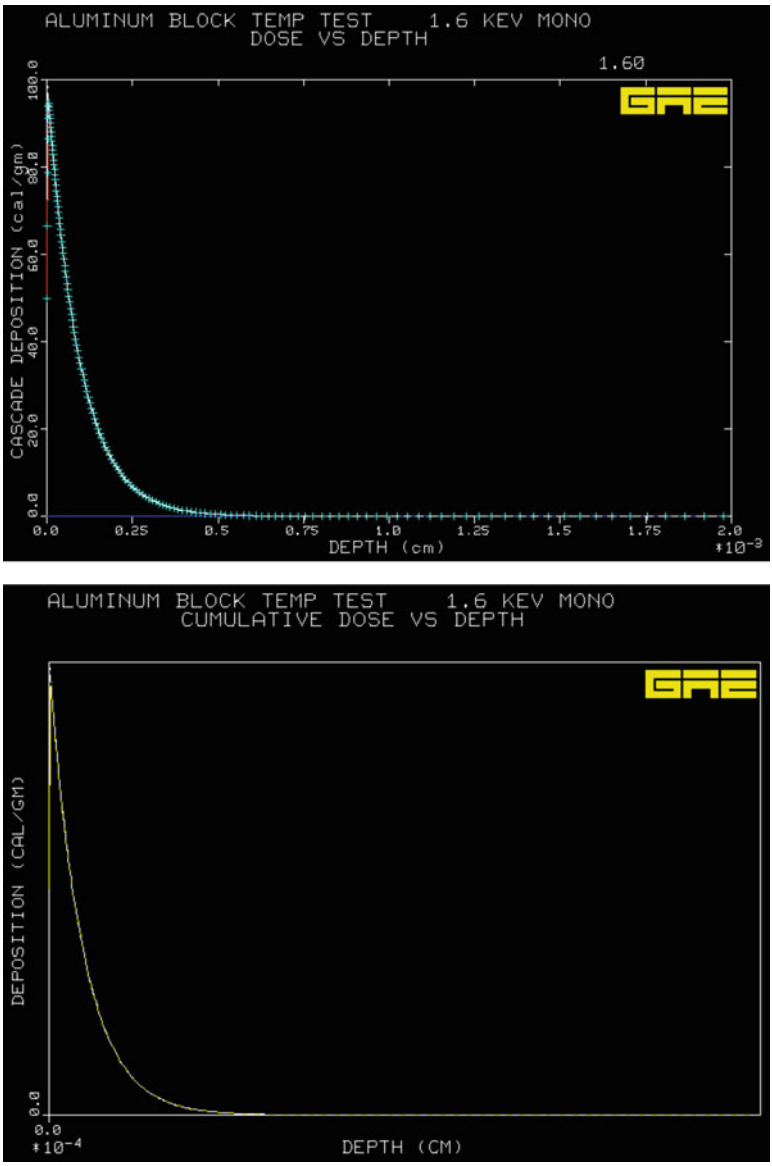


Fig 1.1 Sample output of PUFF74 using Universal Graphics Library from GAE

1.3 PUFF-TFT: A Material Response Computer Code

The PUFF-TFT code has now been updated (version 5.0) to allow modeling of sample responses to sudden energy loading (e.g., X-rays or lasers) for arbitrary starting temperatures. Problems can be run for any initial temperature, both

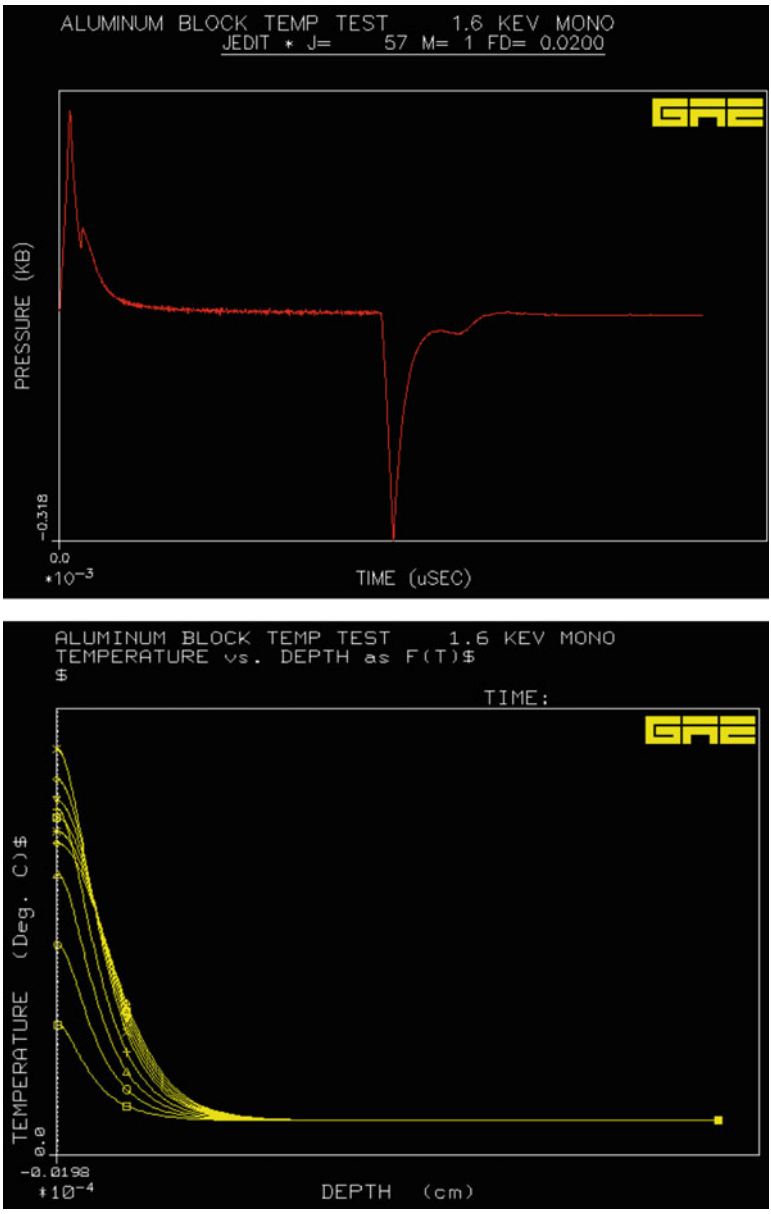


Fig 1.1 (continued)

elevated and, most importantly, for cryogenic conditions. Updates have also been made in the stress response for the "thermal-only" mode, especially for the cool-down stresses after plastic flow. Likewise, the code tracks material properties (yielding, shear module, spall strengths) for cryogenic conditions.

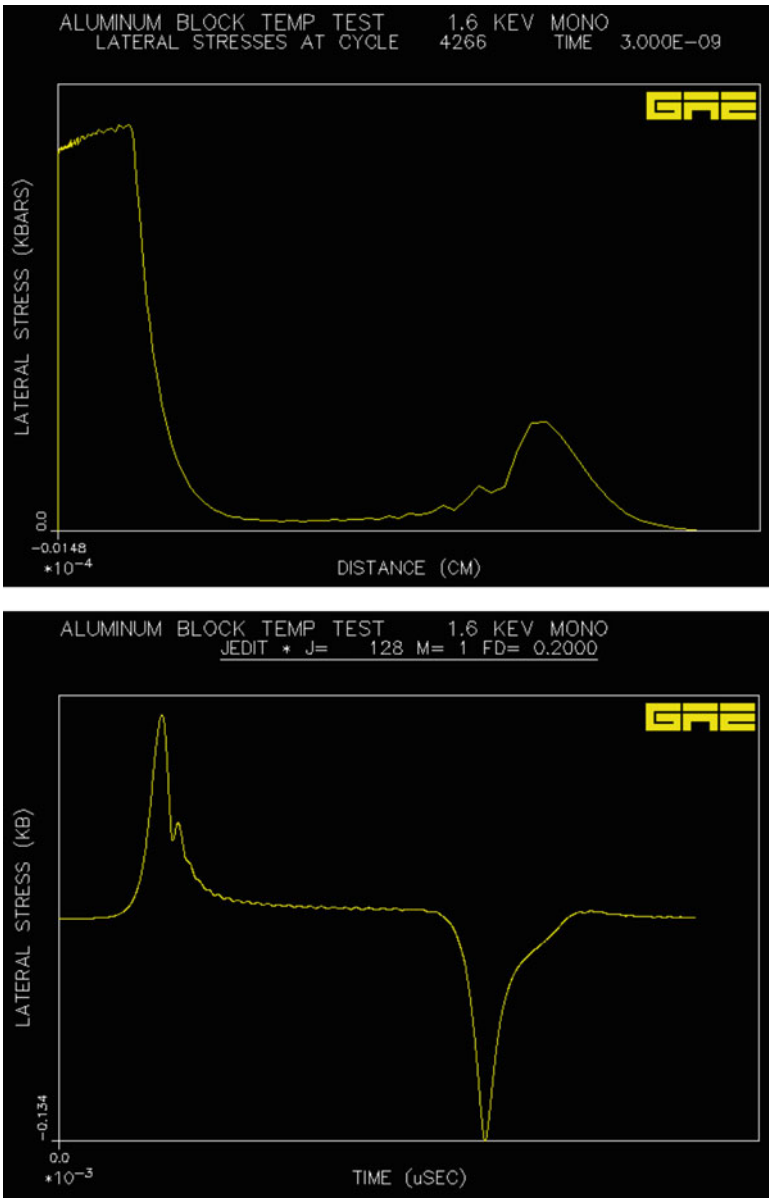


Fig 1.1 (continued)

The code amendments have been done in a "transparent" manner for the user, requiring the minimum of input parameter changes. To activate this, the code maintains the existing convention of:

$$\text{Enthalpy} = 0.0\ \text{cal/g at temperature} = 25\,^{\circ}\text{C}$$

and temperature continues to be in degrees centigrade. Consequently, for that equal to 25 °C, the code will start with a nonzero enthalpy. For $T > 25\,^{\circ}\text{C}$, this initial enthalpy will be positive, whereas for $T < 25\,^{\circ}\text{C}$, the enthalpy is negative.

The previous code version did not distinguish between "dose" (the added energy due to X-rays, thermal flow, etc.) and "enthalpy." This was appropriate, since both terms initialized with a common value of zero. The new code makes the distinction, since dose still starts from zero enthalpy.

The "transparent" amendments are such that the user continues to use the existing database for such parameters as melt energy, vapor energy, and latent heats. Likewise, for $T > 25\,^{\circ}\text{C}$, the existing polynomial coefficients to describe specific heats, enthalpies, and conductivities are maintained.

The code was written for the Air Force Weapon Laboratory (AFWL) primarily to allow evaluation of thin-layer stack response to X-ray deposition resulting in one-dimensional (1-D) strain stress response. The code takes into account the X-ray generation of secondary cascade particles (photoelectrons, Auger electrons, and fluorescent photons) using a cascade routine and incorporates a thermal condition routine allowing the effects of rapid thermal diffusivity to be included.

The output of the X-ray/cascade/thermal routine is used as input to an updated version of the PUFF74 hydrodynamic code, which includes hydrodynamic, elastoplastic, porous, and dispersive material responses in a fully coupled manner, and also accounts for simple phase changes.

The formulation of differential equations follows either Eulerian or Lagrangian descriptions. The Eulerian description is a spatial description; while the Lagrangian is a material description. In an Eulerian framework, all grid points, and consequently cell boundaries, remain fixed with time. Mass, momentum, and energy flow across cell boundaries. In a Lagrangian description, the grid points are attached to the material and move with the material. In this formulation, mass within a cell is invariant, but the volume of the cell may change with time because of expansion or compression, of the materials.

The PUFF-TFT code calculates stress wave formation and propagation by numerical integration of the conservation equations in a one-dimensional Lagrangian coordinate system. The TFT package accounts for the effects of dose enhancement due to the transport of secondary particles with ranges comparable to the thickness of the thin material layers and thermal conduction between thin material layers. These two modifications (among others) more accurately portray the degree of energy sharing between thin layers, thereby modifying the expected energy depositions based on normal X-ray interactions and possibly altering the anticipated thermomechanical response of the medium.

The PUFF74 code, originally developed in the mid-1960s, has undergone a number of revisions to become a flexible material response code that includes the effects of material strength, porosity, and fracture for both homogeneous and composite materials. The code calculates stress wave formation and propagation

by numerical integration of the conservation equations in a one-dimensional Lagrangian coordinate system. In addition to the hydrodynamic equation of state, which is required for all materials, the code contains an elastic–plastic model for strength effects, a P-Alpha porosity model for treating irreversible compaction, and four models for treating strain rate-dependent or dispersive effects.

Below you can see some graphics output sample of PUFF-TFT code (Fig. 1.2).

1.3.1 Availability of PUFF-TFT Computer Code

The Windows/PC version of this code is available from Galaxy Advanced Engineering, Inc. for purchase price. This version has been modified from its original version that used to run on CDC computer, and users need to obtain their own copy by applying to Oak Ridge National Laboratory technology transfer office or contact the Galaxy Advanced Engineering, Inc. for its Windows/PC version. You can find more detail how to obtain and purchase by referring to the following URL: https://www.gaeinc.com.

1.4 SANDYL: A Monte Carlo Three-Dimensional Computer Code

SANDYL is a FORTRAN code for computing, photon–electron transport, and deposition in complex systems by the Monte Carlo method. In this computation, a large number of possible particle trajectories are generated one at a time, and, as the particle proceeds through the material of the system, contributions to the quantities making up the desired information are tallied. After a number of trajectories, the averages of these quantities are statistical approximations, to the solution. All histories of source Andy secondary particles with energies in the 1-keV to 1000-MeV ranges are followed through the system.

The problem geometry is divided into zones, of homogeneous atomic composition bounded, by sections of planes and quadrics. Thus, the material of each zone is a specified element or combination of elements.

For a photon history, the trajectory is generated by following the photon from scattering to scattering using then various probability distributions to find distances between collisions, types of collisions, types of secondary, and their energies and scattering angles. The photon interactions are photoelectric absorption (atomic ionization), coherent scattering, incoherent scattering, and pair production. The secondary photons include bremsstrahlung, fluorescence photons, and $e^+ - e^-$ annihilation radiation.

The condensed-history Monte Carlo method is used for the electron transport. In a history, the spatial steps taken by an electron are precomputed and may include

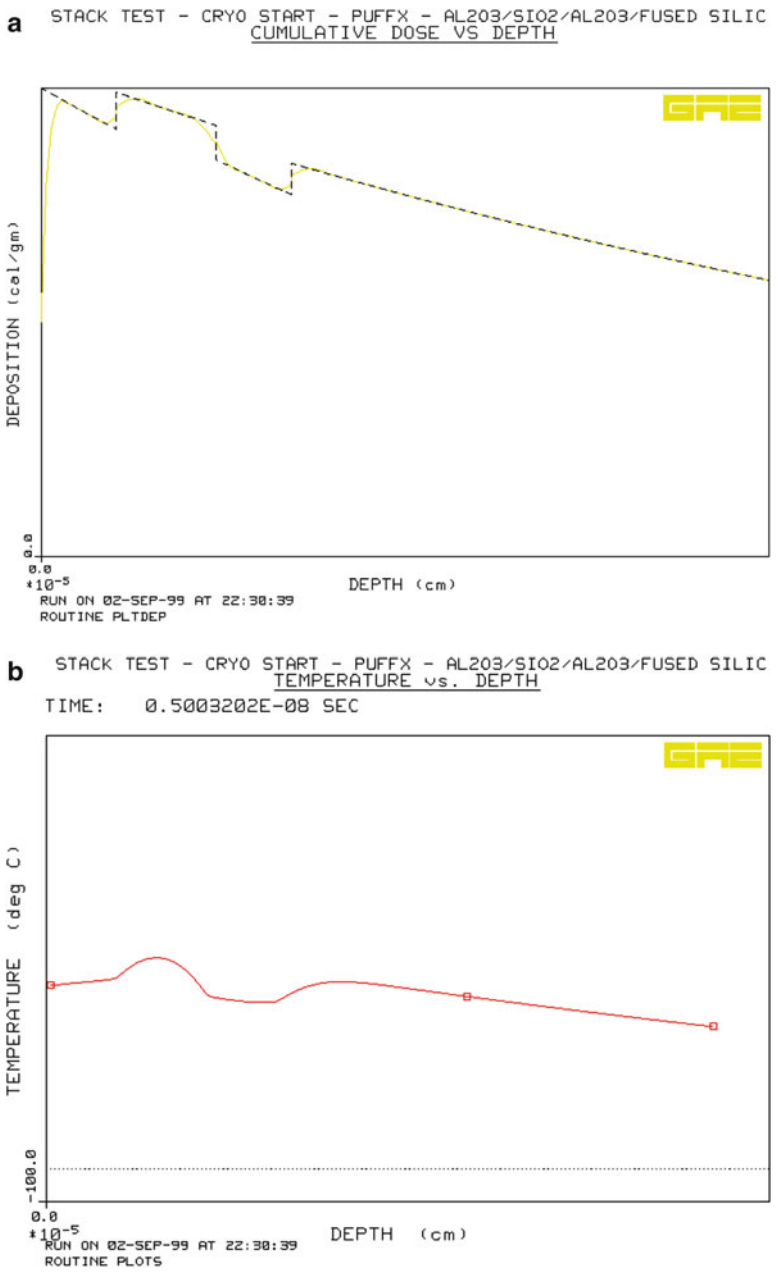


Fig. 1.2 Sample output of PUFF-TFT using Universal Graphics Library from GAE

the effects of a number of collisions. The corresponding scattering angle and energy loss in the step are found from the multiple scattering distributions for these quantities. Atomic ionization and secondary particles are generated within the step according to the probabilities for their occurrence.

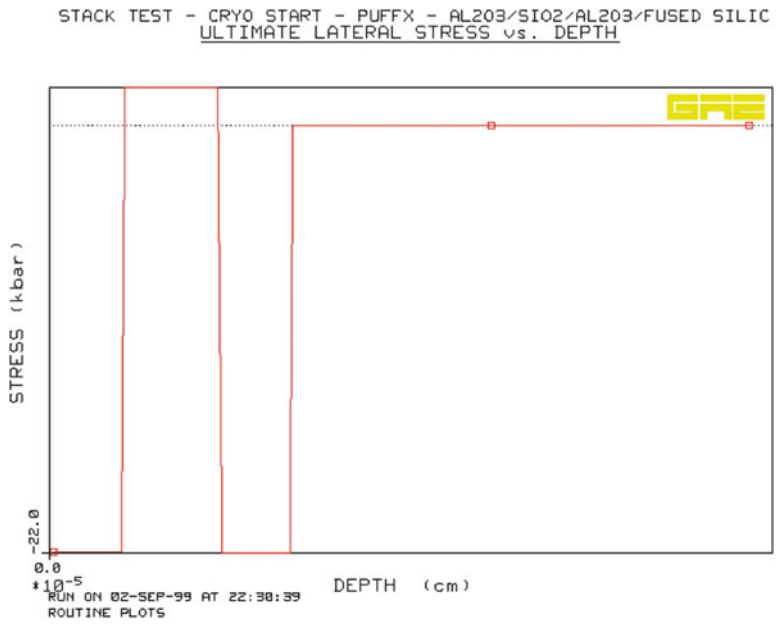


Fig. 1.2 (continued)

Electron energy loss is through inelastic electron–electron collisions, bremsstrahlung generation, and polarization of the medium (density effect). Included in the loss is the fluctuation due to the variation in the number of energy loss collisions in a given Monte Carlo step (straggling). Scattering angular distributions are determined from elastic nuclear-collision cross sections corrected for electron–electron interactions. The secondary electrons include knock-on, pair, Auger (through atomic ionizations), Compton, and photoelectric electrons.

SANDYL is a Monte Carlo three-dimensional code for calculating combined photon–electron transport in complex systems. SANDYL incorporates material from the SORS photon and ETRAN photon–electron codes. Major additions and modifications occur in the atomic ionization and relaxation routines and in the general geometry multiple-material aspects of the electron transport.

SANDYL uses the Monte Carlo method. In its computations, a large number of possible particle trajectories are generated one at a time, and, as the particle proceeds through the material of the system, contributions to the quantities making up the desired information are tallied. After a number of trajectories, the averages of these quantities are statistical approximations to the solution.

The problem geometry is divided into zones of homogeneous atomic composition bounded by sections of planes and quadrics. Thus, the material of each zone is a specified element or combination of elements. For a photon history, the trajectory is generated by following the photon from scattering to scattering using the various probability distributions to find distances between collisions, types of collisions,

types of secondary, and their energies and scattering angles. The condensed-history Monte Carlo method is used for the electron transport.

The code does time- and space-dependent transport calculations of the photon–electron cascade in complex systems. All generations of particles in the 1-keV to 1000-MeV energy range are followed.

1.4.1 Availability of SANDYL Computer Code

The Windows/PC version of this code is available from Galaxy Advanced Engineering, Inc. for purchase price. This version has been modified from its original version that used to run on CDC computer, and users need to obtain their own copy by applying to Oak Ridge National Laboratory technology transfer office or contact the Galaxy Advanced Engineering, Inc. for its Windows/PC version. You can find more detail on how to obtain and purchase by referring to the following URL: https://www.gaeinc.com.

1.5 ASTHMA88 (Axisymmetric Transient Heating and Material Ablation) Code

The ASTHMA88 program has been developed for computing the 2-D symmetric transient thermochemical response of decomposing materials subject to hyperthermal convective and radiative environments. The ASTHMA88 code employs an implicit/explicit, finite difference computational procedure with a fixed two-dimensional grid whose layout is independent of the physical axes. The numerical modeling includes equations for mass and energy conservation and material decomposition; the flow of pyrolysis gas through the porous, decomposing solid; the calculation of material properties as a function of temperature and material state; general ablating surface and back wall/side wall boundary conditions; and a comprehensive surface energy balance which accounts for convection and radiation absorption, reradiation, in-depth condition, surface ablation, pyrolysis gas flow, transpiration effects, and thin-layer mechanical removal or surface melting.

Validation studies demonstrate excellent agreement with other standard thermochemical analysis codes, i.e., CMA (1-D, decomposing) and ASTHMA81 (2-D, non-decomposing).

The **ASTHMA88** code can handle multiple decomposing and non-decomposing, anisotropic materials in simple or complex two-dimensional axisymmetric configurations. Surface boundary conditions may be described in three options:

1. Simple specified temperature and recession
2. Specified heat flux with no recession

General thermochemical model is incorporating both equilibrium and nonequilibrium computations, for any material exposed to any convective and/or radiative environment.

1.5.1 Availability of ASTHMA88 Computer Code

The Windows/PC version of this code is available from Galaxy Advanced Engineering, Inc. for purchase price. This version has been modified from its original version that used to run on VAX/VMS computer, and users need to obtain their own copy from this company. Contact the Galaxy Advanced Engineering, Inc. for its Windows/PC version. To our knowledge the code is no longer available from government agencies or its contractors. You can find more detail on how to obtain and purchase by referring to the following URL: https://www.gaeinc.com.

1.6 ALE3D (Arbitrary Lagrangian/Eulerian Multi-Physics 3D) Computer Code

Composite materials are used in many advanced application systems and structures at Lawrence Livermore National Laboratory (LLNL). We have previously enhanced our ability to simulate structural response and progressive failure of composite systems in ALE3D (an arbitrary Lagrangian/Eulerian multi-physics code developed at LLNL) by porting an existing composite constitutive model (Model 22, the fiber composite with damage model) from DYNA3D (a nonlinear, explicit, 3-D FEM code for solid and structural mechanics). This year, a more advanced model (DYNA3D Model 62, the unidirectional elastoplastic composite model) has been implemented. Experiments were conducted to validate the elastic response of the model and to give insights and data needed for the addition of a failure algorithm into the model.

They implemented the unidirectional elastoplastic composite model into ALE3D. This included implementing the ability to input orthotropic orientation data into prescribed local volume elements. Another modeling goal was to enhance the model by incorporating a failure algorithm that includes matrix delaminating, fiber tensile, and fiber compressive failure. Several experiments were conducted to provide data for the verification and validation of the model's implementation in ALE3D.

The improved fiber composite material models can be used in simulations (to failure) in the many LLNL programs, such as those for composite munitions, armor penetration, pressure vessels, and rocket motors. This project has been beneficial in supporting the composite modeling efforts within the DOD Joint

Munitions Program and the Focused Lethality Munitions Program. This study supports LLNL's engineering core competency in high-rate mechanical deformation simulations of large complex structures by providing an enhanced capability to model composite structures with ALE3D.

The implementation of the fiber composite with damage model into ALE3D, which was completed in the first year of this project, was verified with several code-to-code comparisons. The hoop stresses in pressurized cylinders from simulations run with DYNA3D, with the new fiber composite model in ALE3D, and with an existing anisotropic ALE3D model, all agreed within 1 %. This included both explicit and implicit ALE3D runs.

The unidirectional elastoplastic composite model was implemented into ALE3D. An important part of this task was creating an algorithm to initialize and update material directions at the ply and element levels. The model was validated using the same pressurized-cylinder simulations described above, and the results were found to closely match the DYNA3D predictions.

Composite failure mechanisms can be divided into two types: intra-ply failure mechanisms, such as fiber breakage, matrix failure (cracking/crushing), and fiber buckling and inter-ply failure mechanisms involving ply delaminating.

Intra-ply failure can be applied at the ply level and so fits in well with this model's "unit cell" approach. Inter-ply failure that includes crack opening between plies and plies sliding relative to each other affects all layers simultaneously and so is more difficult to implement. All the relevant mathematical expressions necessary for these functionalities have been derived, and the corresponding changes to the existing code have outlined. Implementation will be undertaken next year.

A series of compression tests to failure were conducted on eight different composite cylinder specimens with different fiber, fiber orientations, and resins. The data collected on the stiffness, Poisson's ratio, and ultimate strength of each specimen provide model validation data for the newly implemented Models 22 and 62. The data also provide an expanded source of failure data for upcoming failure model validation in ALE3D (Fig. 1.3).

Fig. 1.3 Fiber composite compression cylinder with 1.0-in.-diameter pin

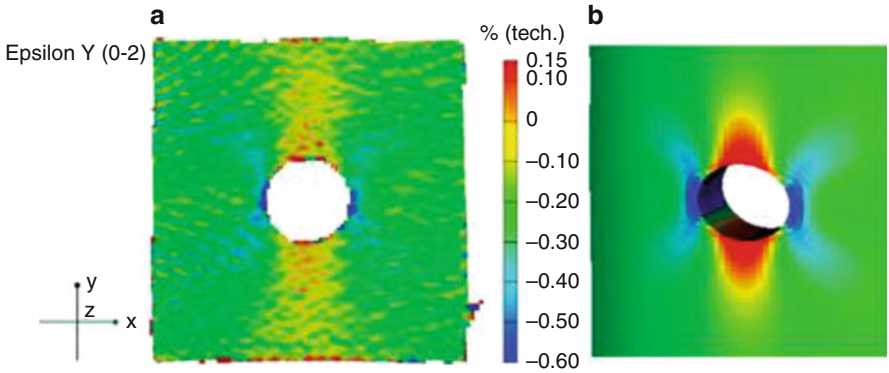


Fig. 1.4 (**a**) Aramis axial stain results for fiber composite compression cylinder with no pin at 300,000 lbs of load. (**b**) The ALE3D simulation

Fig. 1.5 Composite shear specimen from section of a MK82

Strain concentration factors in fiber composite cylinders with holes and bonded pins were measured using the Aramis video strain measurement system. The basic fiber composite cylinder with pin configuration is shown in Fig. 3.62. Figure 3.63 shows a comparison for the case of no pin (open hole) between the measured experimental data and the simulated response from ALE3D. The results appear to be very similar.

Strain concentration factors due to focused shear in composites were measured using the specimen shown in Fig. 1.4. This sample was loaded in compression to produce a concentrated shear band in the composite sample. The Aramis load strain curve is shown in Fig. 3.65 (Fig. 1.5).

In a proposed follow-on project, we will continue to improve fiber composite modeling in ALE3D, with an emphasis on local bending response and progressive damage. We plan to implement ply-level capabilities and damage algorithms taken from a specialized LLNL ply-level composite code known as ORTHO3D and verify their implementation experimentally (Fig. 1.6).

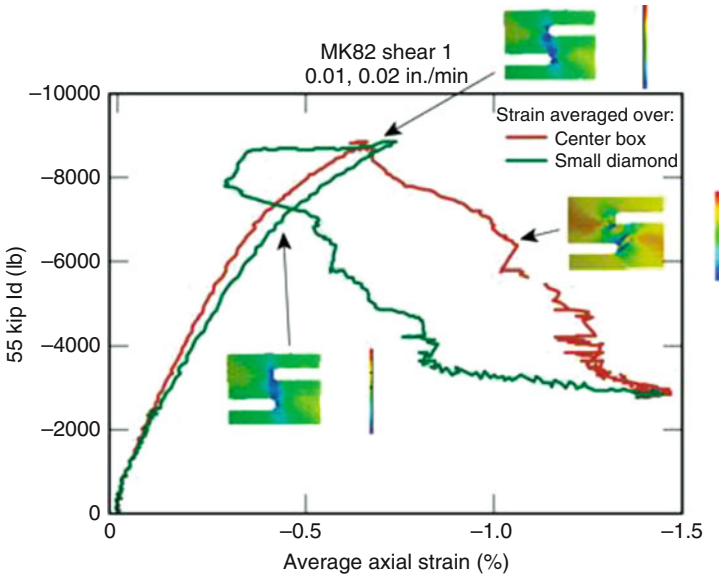


Fig. 1.6 Shear strain concentration in composite Mk82 shear specimen

1.6.1 ALE3D Program Availability

This code is available from Lawrence Livermore National Laboratory, and users need to obtain their own copy by applying to LLNL technology transfer office or contact the author of the code Andrew Anderson (925) 423-9634 or refer to the site of this code at the following URL: https://www-eng.llnl.gov/mod_sim/mod_sim_tools.html.

1.7 CTH Computer Code

CTH is a multi-material, large deformation, strong shock wave, solid mechanics code that runs on most UNIX workstations and MPP supercomputers. CTH is one of the most heavily used computational structural mechanics codes on DOD high-performance computing (HPC) platforms. While CTH includes some internal graphics capabilities, it is preferable to take advantage of widely used scientific visualization packages like EnSight and ParaView to analyze the results of calculations. A new method has been devised that extends the capabilities of CTH to allow three-dimensional polygonal models to be written directly from a running calculation in a format compatible to both EnSight and ParaView. Additionally, an interpreter for the scripting language Python has been embedded into CTH and its post-processor Spymaster. Embedded Python allows for almost limitless, parallel

capabilities to be added that do not require a recompilation or relinking of the CTH executable. Examples of these capabilities include one- and two-way code coupling and behind armor debris (BAD) applications.

The latest version of the widely used shock wave physics computer code, CTH, developed by Sandia National Laboratories, will soon be available to customers nationwide. The code simulates high-speed impact and penetration phenomena involving a variety of materials.

Interest in the new version of the software is particularly high among customers like the Department of Energy (DOE) and Department of Defense (DOD), which use the software for studying weapon effects, armor/anti-armor interactions, warhead design, high-explosive initiation physics, and weapon safety issues. Major users include the national laboratories; the army, navy, and air force laboratories; and their subcontractors. At Sandia, a DOE laboratory, the code is used in national missile defense, hazardous material dispersal by explosive detonation, weapon components design, and reactive materials research (Fig. 1.7).

For armor/anti-armor design—of interest to DOD—the software allows users to determine which types of bullets or projectiles can best penetrate armor. It also provides information about how to design an improved penetration protection mechanism.

"This new version is really exciting because it offers a computational capability never before available in this type of code, an adaptive mesh refinement model [AMR]," says Paul Taylor, head of the CTH project at Sandia. "AMR gives the software the ability to increase resolution and accuracy in those regions of a simulation where it is needed and reduce resolution in those regions where it is not. For example, in the simulation of a projectile penetrating a target material, greater resolution can be achieved in the region surrounding the impact interface between the two materials where large distortions and high strain rates are occurring."

The medical community is also paying attention to Sandia's CTH software. Taylor currently has a small collaborative research effort underway with the University of New Mexico School of Medicine, which is interested in using the shock physics code to better understand brain injury caused by physical trauma, such as a person's head hitting a car windshield. Using the magnetic resonance imaging (MRI) of an individual's head to construct a CTH model, simulations can be performed showing how shock waves travel through the head and cause brain damage.

The software breaks down the penetration simulation into millions of grid-like "cells." As the modeled projectile (such as a copper ball impacting a steel plate) impacts and penetrates the target, progressively smaller blocks of cells are placed around the projectile, each showing in detail the deformation and breakup of the ball and target plate.

CTH with the AMR enhancement also offers the ability to analyze problems involving sophisticated materials with greater accuracy. With the addition of new material models, it can simulate a wider variety of materials, including metals, ceramics, plastics, composites, high explosives, rocket propellants, and gases.

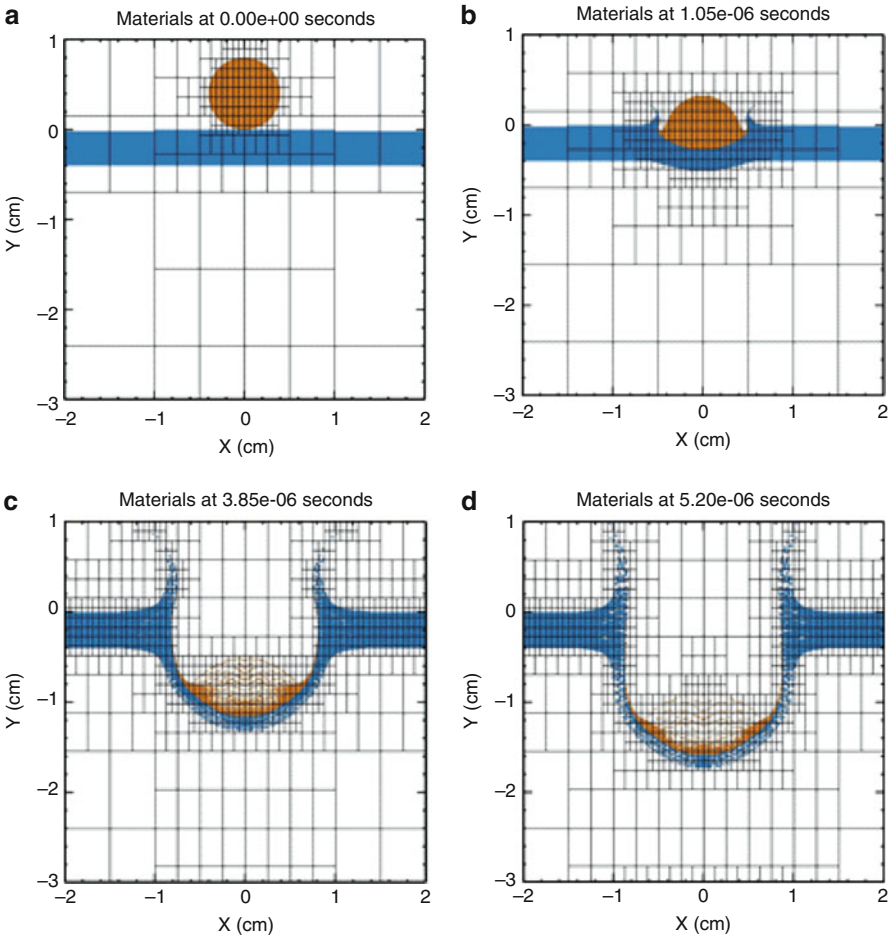


Fig. 1.7 (**a**) CTH run time for $t = 0.00\text{e}+00$ s. (**b**) CTH run time for $t = 1.05\text{e}+06$ s. (**c**) CTH run time for $t = 3.85\text{e}+06$ s. (**d**) CTH run time for $t = 5.20\text{e}+06$ s

Sandia developed the early precursor to CTH in the 1970s for one-dimensional problems, expanding it to simulate problems in two and three dimensions in the 1980s.

The labs began licensing the shockwave physics code in the early 1990s to DOE, DOD, their contractors, and some private US companies with interests in shock physics.

An updated version of the software, which is export controlled, is distributed to customers about every 18 months. Currently 259 licenses have been issued.

DOD, DOE, and their contractors receive use licenses for a small distribution fee. Commercial companies can purchase licenses for $25,000. The updated software will be distributed on CDs at a cost of $400 for each noncommercial, licensed customer.

One of the most appealing aspects of CTH for users is that it can run on almost any computer platform. Taylor offers CTH classes at Sandia several times a year to users from all over the country.

1.7.1 Availability of CTH Computer Code

This code is available from Sandia National Laboratory. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US Department of Energy under contract DE-AC04-94AL85000. With main facilities in Albuquerque, NM, and Livermore, CA, Sandia has major research and development responsibilities in national security, energy and environmental technologies, and economic competitiveness.

Media contact:
Chris Burroughs, coburro@sandia.gov, (505) 844-0948

Technical contact:
Paul Taylor, pataylo@sandia.gov (505) 844-1960

1.8 HYPUF, Stress Wave Response Computer Code

HYPUF is a stress wave response code that has the ability to calculate ionization effects in high-temperature, high-density plasmas. As such, HYPUF/PC is a derivative of the PUFF-66 code. HYPUF is also a code for any defense contractor having a need to calculate the response of materials to radiation induced stress waves.

The modification to present **HYPUF** code available in PC program is part of a continuing program to provide a code suitable for analysis of material interaction with X-ray lasers and other high-intensity radiation sources. Previous version of this code included automatic zoning, rezoning, and spall (fracture) capabilities. The modifications in the present code include elastic–viscoplastic.

Maxwell dispersion and Bade geometric dispersion material response models are implemented, restructuring of the code to facilitate future modifications and numerous minor corrections to the equation of state and ionization equation of state subroutines. All the above three models are incorporated as closely as possible to the way they were implemented in PUFF74 code. The only differences between the implementation in the two codes were that imposed by the fact that HYPUF/PC is a temperature-based rather than energy-based code and that **HYPUF/PC** has its equation of state package completely separate from the HYDRO routine.

The elastic–viscoplastic model is an extension of the elastic–plastic model, which is used to calculate stress deviators in solid materials. In the elastic–viscoplastic model, the stress deviator can overshoot the yield surface value. The stress deviator is computed incrementally from the differential equation.

1.8.1 Availability of HYPUF, Stress Wave Response Computer Code

The Windows/PC version of this code is available from Galaxy Advanced Engineering, Inc. for purchase price. This version has been modified from its original version that used to run on VAX/VMS computer, and users need to obtain their own copy from this company. Contact the Galaxy Advanced Engineering, Inc. for its Windows/PC version. To our knowledge the code is no longer available from government agencies or its contractors. You can find more detail on how to obtain and purchase by referring to the following URL: https://www.gaeinc.com.

1.9 DYNA2D and DYNA3D Computer Codes Series

There is a series of computer codes that were released by Lawrence Livermore National Laboratory such as DYNA2D and DYNA3D, as Lagrangian finite element methods of analysis. DYNA2D and DYNA3D are an explicit finite element code for analyzing the transient dynamic response of three-dimensional solids and structures. The element formulations available include one-dimensional truss and beam elements, two-dimensional quadrilateral and triangular shell elements, two-dimensional delamination and cohesive interface elements, and three-dimensional continuum elements.

Many material models are available to represent a wide range of material behavior, including elasticity, plasticity, composites, thermal effects, and rate dependence. In addition, DYNA2D and DYNA3D have a sophisticated contact interface capability, including frictional sliding and single-surface contact, to handle arbitrary mechanical interactions between independent bodies or between two portions of one body. Also, all element types support rigid materials for modeling rigid body dynamics or for accurately representing the geometry and mass distribution of a complex body at minimum cost. A material model driver with interactive graphics display is integrated into DYNA2D and DYNA3D to allow computation of the stress response to any prescribed strain history without inertial effects. This feature allows accurate assessment of the representation of complex material behavior by the numerical constitutive model in DYNA2D and DYNA3D.

The 3-D version of these codes, DYNA3D, is an explicit, three-dimensional, finite element program for analyzing the large deformation dynamic response of inelastic solids and structures. DYNA3D contains 30 material models and 10 equations of state (EOS) to cover a wide range of material behavior. The material models implemented are: elastic, orthotropic elastic, kinematic/isotropic plasticity, thermoelastoplastic, soil and crushable foam, linear viscoelastic, Blatz–Ko rubber, high-explosive burn, hydrodynamic without deviatory stresses, elastoplastic hydrodynamic, temperature-dependent elastoplastic, isotropic elastoplastic, isotropic elastoplastic with failure, soil and crushable foam with failure, Johnson–Cook

plasticity model, pseudo TENSOR geological model, elastoplastic with fracture, power law isotropic plasticity, strain rate-dependent plasticity, rigid, thermal orthotropic, composite damage model, thermal orthotropic with 12 curves, piecewise linear isotropic plasticity, inviscid two-invariant geologic cap, orthotropic crushable model, Mooney–Rivlin rubber, resultant plasticity, closed form update shell plasticity, and Frazer–Nash rubber model. The IBM 3090 version does not contain the last two models mentioned.

The hydrodynamic material models determine only the deviatoric stresses. Pressure is determined by one of ten equations of state including linear polynomial, JWL high explosive, Sack "Tuesday" high explosive, Gruneisen, ratio of polynomials, linear polynomial with energy deposition, ignition and growth of reaction in HE, tabulated compaction, and tabulated and TENSOR pore collapse. DYNA3D generates three binary output databases. One contains information for complete states at infrequent intervals; 50–100 states are typical. The second contains information for a subset of nodes and elements at frequent intervals; 1000–10,000 states are typical. The last contains interface data for contact surfaces.

The method of solution is based on a contact–impact algorithm permit gaps sliding along material interfaces with friction. All versions except for the IBM3090 include an interface type defining one-way treatment of sliding with voids and friction. By a specialization of this algorithm, such interfaces can be rigidly tied to admit variable zoning with no need for transition regions. Spatial discretization is achieved by implementation of Hughes–Liu rectangular beams and shells, Belytschko–Tsay shells and beams, triangular shell elements based on work by Belytschko and colleagues, and 8-node solid-shell elements. All element classes can be included as parts of a rigid body. Three-dimensional plane-stress constitutive subroutines update the stress tensor for the shell elements such that the stress component normal to the shell mid-surface is zero. One constitutive evaluation is made for each integration point through the shell thickness. The 8-node solid element uses either one-point integration or the Flanagan and Belytschko constant stress formulation with exact volume integration. Zero energy modes in the shell and solid elements are controlled by either an hourglass viscosity of stiffness. The equations of motion are integrated in time by the central difference method. A Jaumann stress rate formulation is used with the exception of the orthotropic elastic and the rubber material subroutines which use Green–St. Venant strains to compute second Piola–Kirchhoff stresses which transform to Cauchy stresses.

1.9.1 Availability of DYNA2D and DYNA3D Computer Codes

Both these codes are available from the following site but have no technical support nor are up-to-date with any present computing operating system, and codes have a lot of logical errors that need to be fixed, and based on this author's experience

working with these codes for a long time, they are not easy to fix unless you are willing to spend hours and hours of debugging.

This package is distributed by:
Energy Science and Technology Software Center
P.O. Box 62
1 Science.Gov Way
Oak Ridge, TN 37831
(865) 576-2606 TEL
(865) 576-6436 FAX
E-mail: ESTSC@osti.gov

The Windows/PC versions of these codes are available from Galaxy Advanced Engineering, Inc. [1] Just refer to www.gaeinc.com.

1.10 NIKE2D and NIKE3D Computer Codes Series

NIKE2D is an implicit finite element code for analyzing the finite deformation, quasistatic, and dynamic response of two-dimensional, axisymmetric, plane-strain, and plane-stress solids. The finite element formulation accounts for both material and geometric nonlinearities. A number of material models are incorporated to simulate a wide range of material behavior including elastoplasticity, anisotropy, creep, thermal effects, and rate dependence. Arbitrary contact between independent bodies is handled by a variety of slide-line algorithms. These algorithms model gaps and sliding along material interfaces, including interface friction and single-surface contact. Interactive graphics and rezoning are included for analyses with large mesh distortions. NIKE2D is no longer funded for active development by LLNL or direct user support and is made available on an “as-is” basis. Select hardware projects have chosen to fund limited development or maintenance activities.

NIKE3D is a fully implicit, three-dimensional finite element code for analyzing the finite strain static and dynamic response of inelastic solids, shells, and beams. Spatial discretization is achieved by the use of eight-node solid elements, two-node truss and beam elements, and four-node membrane and shell elements. Over 20 constitutive models are available for representing a wide range of elastic, plastic, viscous, and thermally dependent material behavior. Contact–impact algorithms permit gaps, frictional sliding, and mesh discontinuities along material interfaces. Several nonlinear solution strategies are available, including full-, modified-, and quasi-Newton methods. The resulting system of simultaneous linear equations is either solved iteratively by an element-by-element method or directly by a factorization method, for which case bandwidth minimization is optional. Data may be stored either in or out of core memory to allow for large analyses.

1.10.1 Availability NIKE2D and NIKE3D Computer Codes Series

Both these codes are available from the following site but have no technical support nor are up-to-date with any present computing operating system, and codes have a lot of logical errors that need to be fixed, and based on this author's experience working with these codes for a long time, they are not easy to fix unless you are willing to spend hours and hours of debugging.

This package is distributed by:
Energy Science and Technology Software Center
P.O. Box 62
1 Science.Gov Way
Oak Ridge, TN 37831
(865) 576-2606 TEL
(865) 576-6436 FAX
E-mail: ESTSC@osti.gov

The Windows/PC versions of these codes are available from Galaxy Advanced Engineering, Inc. Just refer to www.gaeinc.com.

1.11 TOPAZ2D and TOPAZ3D Computer Codes Series

TOPAZ2D is a two-dimensional, implicit, finite element computer code for heat transfer analysis. It can be used to solve for the steady-state or transient temperature field on two-dimensional geometries. TOPAZ2D is no longer funded for active development or direct user support and is made available on an "as-is" basis. Select hardware projects have chosen to fund limited development or maintenance activities.

TOPAZ3D is a three-dimensional, implicit, finite element computer code for heat transfer analysis. It can be used to solve for the steady-state or transient temperature field on three-dimensional geometries. Material properties may be temperature dependent and either isotropic or orthotropic. A variety of time- and temperature-dependent boundary conditions can be specified, including temperature, flux, convection, and radiation. By implementing the user subroutine feature, users can model chemical reaction kinetics and allow for any type of functional representation of boundary conditions and internal heat generation. TOPAZ3D can solve problems of diffuse and specular band radiation in an enclosure coupled with conduction in the material surrounding the enclosure. Additional features include thermal contact resistance across an interface, bulk fluids, phase change, and energy balances. Thermal stresses can be calculated using the solid mechanics code NIKE3D, which reads the temperature state data calculated by TOPAZ3D (Fig. 1.8).

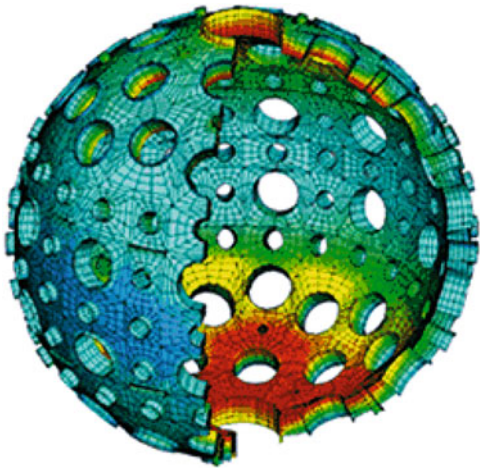


Fig. 1.8 TOPAZ3D and NIKE3D output

The TOPAZ3D and NIKE3D codes are used to analyze the expansion of the National Ignition Facility's laser target chamber resulting from the heat of a laser shot and its contraction to equilibrium in the cool-down period

1.11.1 Availability TOPAZ2D and TOPAZ3D Computer Codes Series

Both these codes are available from the following site but have no technical support nor are up to date with any present computing operating system, and codes have a lot of logical errors that need to be fixed, and based on this author's experience working with these codes for a long time, they are not easy to fix unless you are willing to spend hours and hours of debugging.

This package is distributed by:
Energy Science and Technology Software Center
P.O. Box 62
1 Science.Gov Way
Oak Ridge, TN 37831
(865) 576-2606 TEL
(865) 576-6436 FAX
E-mail: ESTSC@osti.gov

The Windows/PC versions of these codes are available from Galaxy Advanced Engineering, Inc. Just refer to www.gaeinc.com.

References

1. Duffner RW (1997) Airborne laser: bullets of light. Plenum Publishing Corporation, New York
2. Anderberg B, Wolbarsht ML (1992) Laser weapons: the dawn of a new military age. Plenum Publishing Corporation, New York
3. Cooper AF (1969) Automatic initial flesh distribution f o r PUFF 66, AFWL-TR-69-63. Kirtland AFB, New Mexico

Chapter 2
Laser Technology

The development of lasers has been an exciting chapter in the history of science and engineering. It has produced a new device with potential for applications in an extraordinary variety of fields. Einstein developed the concept of stimulated emission on theoretical grounds. Stimulated emission is the phenomenon that is utilized in lasers. Stimulated emission produces amplification of light so that buildup of high-intensity light in the laser can occur. Einstein described the fundamental nature of the stimulated emission process theoretically.

This characterization of stimulated emission did not lead immediately to the laser. Additional preliminary work on optical spectroscopy was done in the 1930s. Most of the atomic and molecular energy levels that are used in lasers were studied and investigated during those decades.

2.1 Basic Principles

The word laser is an acronym for light amplification by stimulated emission of radiation, although common usage today is to use the word as a noun—laser—rather than as an acronym—LASER.

A laser is a device that creates and amplifies a narrow, intense beam of coherent light.

Atoms emit radiation. We see it every day when the "excited" neon atoms in a neon sign emit light. Normally, they radiate their light in random directions at random times. The result is incoherent light—a technical term for what you would consider a jumble of photons going in all directions.

The trick in generating coherent light—of a single or just a few frequencies going in one precise direction—is to find the right atoms with the right internal storage mechanisms and create an environment in which they can all cooperate—to give up their light at the right time and all in the same direction.

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_2

In a laser, the atoms or molecules of a crystal, such as ruby or garnet—or of a gas, liquid, or other substance—are excited in what is called the *laser cavity* so that more of them are at higher energy levels than are at lower energy levels. Reflective surfaces at both ends of the cavity permit energy to reflect back and forth, building up in each passage.

Only three basic components are necessary for laser action: a lasing medium, a pumping system that supplies energy to the lasing medium, and a resonant optional cavity. Lenses, mirrors, shutters, saturable absorbers, and other accessories may be added to the system to obtain more power, shorter pulses, or special beam shapes.

2.2 Overall Theme

This report deals with the effects of directed energy weapons, treating such diverse types of weaponry in particular laser and in our case airborne laser (ABL). Although when we talk about directed energy weapon, we can consider such weapon as particle beams, microwaves [1], and even bullets as part of directed energy weapon (DEW) system. In order to understand these weapons and their effects, it is necessary first to develop a common framework for their analysis, and in our particular case, we expand our concentration on just laser as DEW in particular ABL under the scope of this project and related issues of laser range safety tool (LRST).

It is a thesis of this report that all laser weapons (continuous or pulse) may be understood as devices which deposit energy in targets and that the energy which must be deposited to achieve a given level of damage is relatively intensive to the type of laser weapon employed, type of engagement environment, dual time on target, and type of targets these weapons are engaged.

Of course, energy cannot be deposited in a target unless it has first delivered there. Therefore, an important element in understanding laser weapons is knowledge of how they deliver (or "propagate") their energy. Some loss of energy is invariably associated with this propagation, whether it is the atmospheric effect such as a known phenomenon as thermal blooming [1] or delivery system as well as other related technical and obstacle issues. A laser weapon must therefore produce more energy than needed to damage a target, since some of its energy will be lost in propagation. As a result, weapon design depends upon two factors: first, the anticipated target, which determines the energy required for damage, and, second, the anticipated scenario (range, engagement time, etc.) which determines how much energy should be produced to insure that an adequate amount is delivered in the time available and dual time on target.

2.3 A Word About Units

Since our goal is to reduce the jargon associated with different types of laser weaponry to common units, the choice for these common units is obviously of interest. For the most part, we will use metric units of MKS, where length is in meter, mass is in kilograms, and time is in seconds. In these units, energy is expressed as *joules*.

2.4 Developing Damage Criteria

If we are to determine how much energy a weapon must produce to damage a target, we need to know two things:

1. How much energy it takes to damage a target.
2. What fraction of the energy generated will be lost in propagation to it.

These will be developed in detail for different weapon types in subsequent sections. For the moment, we will consider some of the fundamental issues, which affect damage, and propagation of laser weapon independent of its type (CW or pulse).

2.5 The Energy Required for Damage

In order to be quantitative about the amount of energy necessary for damage, we must first define what we mean by damage. For a military system, this could be anything from an upset in a target's computer (in case of microwave weapon) and preventing it from operating or to total vaporization (in the case of laser weapon). These two extremes are usually referred to as "soft" and "hard" damage or kill, respectively. The study of soft damage clearly is beyond the scope of this project and is much more sensitive to specific details of the target system and its shielding as well as related countermeasure under attack than hard damage. Few good references are published and are available for soft kill or damage study [2].

Without knowing the details of a computer, its circuitry designs, and the hardness of its chips and electronic components, we will not know if it has been upset until we see it in operation, whereas vaporizing it produces immediate feedback on the effectiveness of an attack, and it is the subject of this project. On the other hand, vaporizing a target will require more energy than degrading its performance. We will concentrate in this book on hard or catastrophic damage for two reasons: it avoids target-specific details, which are often classified, and it provides a useful first cut at separating weapon parameters, which will almost

certainly result in damage from those for which the likelihood of damage is questionable or for which more detailed analysis is required.

As a simple example of the kind of energies necessary to achieve damage, let us first consider what it takes to vaporize a given target using laser techniques employed in effects of high-power laser radiation [3] or effect of laser radiation on absorbing condensed matter [4].

2.6 The Laser Beam

The laser beam has many unique qualities, which can be manipulated in many ways by the use of different accessories that are added to the basic laser. The beam is characterized by its collimation, coherence, monochromaticity, speed, and intensity [5]. The laser beam is the source of light that can have all the above properties, while the other source of light may possess these properties but not all at the same time.

Collimation in a laser can be very high, which means that the radiation emitted by most lasers is confined to a very narrow beam, which slowly diverges as the beam moves away from the laser source, a phenomena that is known as diffraction.

Diffraction refers to the spreading, or divergence, of light which emerges from an aperture of given diameter, as shown in Fig. 2.1 [6].

In this figure, a beam of light of essentially infinite beam width is passed through an aperture of diameter D. Calculating the beam divergence or diffraction is a matter of elementary geometry analysis, and it can be shown that the angle of divergence θ is related to D and the wavelength, λ, of the beam by relationship $\theta \approx \lambda/D$. The divergence of the beam is normally a small enough angle so that the approximation holds that the sine and tangent of the divergence angle have the same value, with the angle itself expressed in milliradians (a milliradian divergence would mean that a beam would be 1 yard wide at 1000 yards range, 2 yards wide at 2000 yards, and so forth) [5]. Due to nature of wave light, it is impossible to make a laser weapon that is 100 % collimated beam and has no divergence at all (see

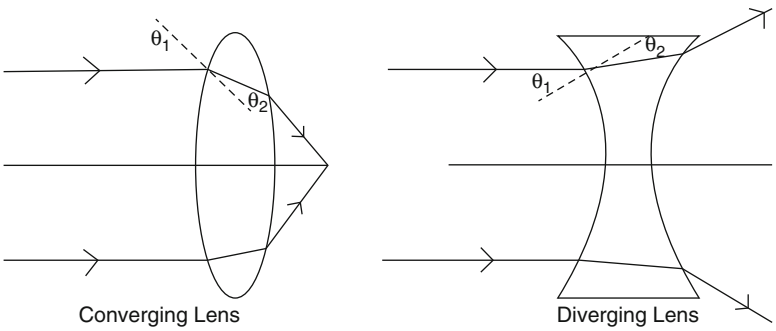


Fig. 2.1 Converging and diverging lenses (Figure has been adapted from Figure 3.13 in Eugene Hecht and Alfred Zajac, Optics, Addison-Wesley, 1976)

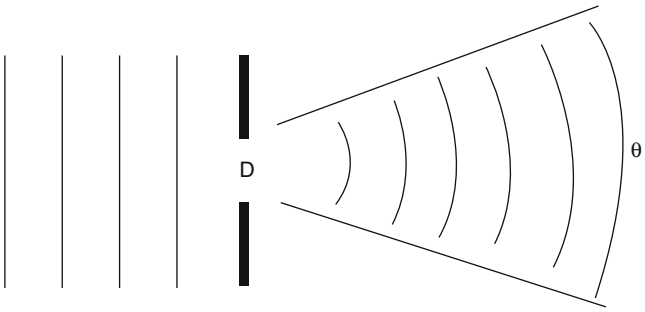


Fig. 2.2 Diffraction of light passing through an aperture

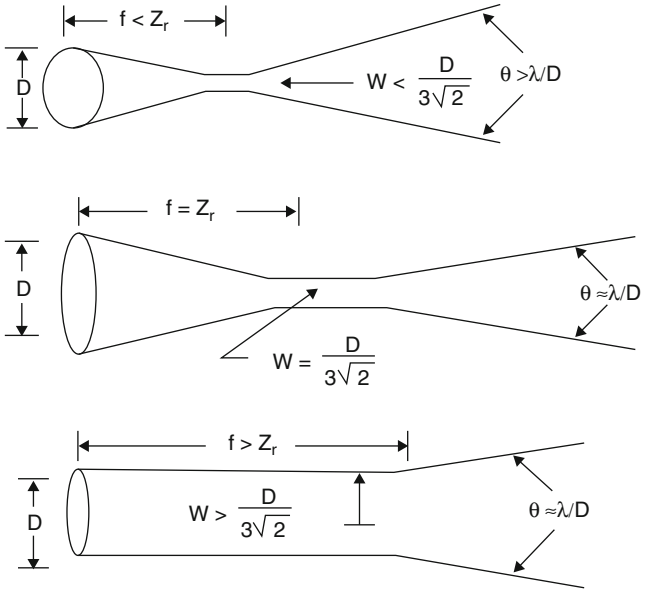


Fig. 2.3 Focusing of a beam of light and the Rayleigh range

Fig. 2.2). However, the angle of divergence of a laser beam can be forced to be as small as possible by usage of a converging lens that is placed in path of the beam. Such approach reduces the effect of divergence to achieve a longer effective beam as illustrated in Fig. 2.3.

These lenses serve as an apparatus to bend the laser beam inward, focusing it to a spot of radius W. The width of the focal spot depends upon the focal length, f, of lens. For shorter length of f, the beam focuses to smaller spot and will diverge rapidly beyond that spot, and for longer, f, the light diverges as it leaves the source of the light at the aperture.

The best optimum focal length at which beam has the best optimum collimation for greatest distance is known as the *Rayleigh range*, Z_r. The beam radius at the

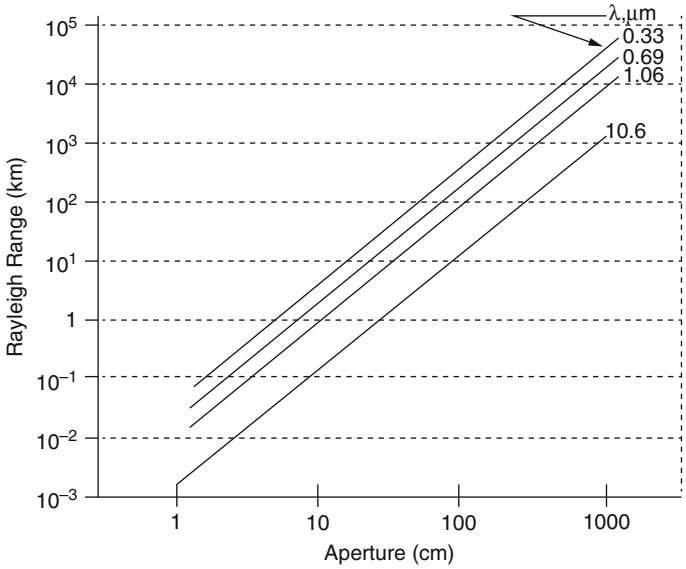


Fig. 2.4 Rayleigh range vs. aperture and laser wavelength

Rayleigh range is $W = D/3\sqrt{2}$, and the Rayleigh range is given by $Z_r = \pi W^2/\lambda$. Therefore, in practical application, laser light can be used as a collimated beam over a distance of about twice the Rayleigh range or about D^2/λ, where D is the aperture from which the light emerges from the weapon and λ the wavelength of the light. If the laser designer wants this laser beam to be focused as much as possible on small spot at long distances, the reciprocal relationship between divergence and the size of the output optics is used (see Fig. 2.4). When a beam with a very small divergence is required, large lenses must be used on the output of laser aperture. Beyond this distance, divergence and diffraction at an angle of about λ/D must be taken into account in evaluating the energy density on target. With ordinary lenses, the focal spot may not be smaller than a few times the wavelength of light. For most military purpose, this is certainly more than sufficient. In some high-energy laser (HEL) weapon systems, a concave mirror is used to focus as much energy on the target as possible.

Note: Propagation within Rayleigh range is known as "near-field" propagation and a greater distance as "far-field" propagation.

Laser can operate in the continuous-wave (CW) or the pulsed mode. The mode of operation depends on whether the pump energy is CW or pulsed. A CW mode laser emits light steadily as long as it is turned on. A pulsed mode laser can have either one single pulse or repeated pulses, possibly on a regular basis in a train. The pulse repetition frequency (PRF) is the number of pulses a laser produces in a given time. The duration of the pulse (or pulse width) and the PRF may vary immensely between different lasers. Lasers are available with a PRF as high as several hundreds of thousands or millions of pulses per second. In a visible beam band, the human eye will not see such a pulsation, and the beam will appear to be CW [5].

One of the most important factors to a designer and user of laser weapons is the energy level delivered by laser beam. Energy is the power emitted by a laser within a given time. The following equation can be used to calculate the intensity of the beam:

$$E = P \cdot t$$

where E is the energy in joules, P is power in watts, and t is time in seconds. The energy of repetitively pulsed lasers is calculated using the average power level emitted over a standard interval, which is usually 1 s.

A high-energy laser weapon designed to down aircraft, missile from several miles away may have several megawatts of power, while a low-energy helium–neon laser such as is used in a lecture hall pointer or a supermarket scanner usually has only a milliwatt or less of average CW power, although the CW power of a helium–neon laser can be as much as 50 mW [5].

2.7 Summary

Present-day laser technology is very extensive and diversified, and, within certain limits, it allows for many civilian as well as military applications. Military staff, defense researcher institutes, and defense industries are constantly looking for new laser concept that are suitable for military application and that will fulfill the very tough but realistic battlefield requirements. Many new military laser systems will most certainly be designed to back up their military needs. Thus, if and when realistic battlefield laser weapons concepts pass through the research and development phase, there will be a strong laser industry already in existence to mass-produce these weapons [5].

References

1. Volkovitsky OA, Sedunov YS, Semenov LP (1992) Propagation of intensive laser radiation in clouds, vol 138. AIAA, Washington, DC
2. Meassenger GC, Ash MS (1992) The effects of radiation on electronic systems, 2nd edn. VNB, New York
3. Ready JF (1971) Effects of high-power laser radiation. Academic, New York, NY
4. Fedorov VB (1990) Effect of laser radiation on absorbing condensed matter (trans: Moore R). In: Prokhorov AM (ed) Proceeding of the Institute of General Physics Academy of Science of USSR, vol 13. Nova Science, Commack, New York
5. Anderberg B, Wolbarsht ML (1992) Laser weapons: the dawn of a new military age. Plenum Publishing Corporation, New York
6. Nielson PE. Effects of directed energy weapons sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjS4qKu24LOAhVR7GMKHc0MBBgQFggfMAA&url=http%3A%2F%2Fwww.dtic.mil%2Fcgibin%2FGetTRDoc%3FAD%3DADA476195&usg=AFQjCNGbehu8jdWvh9ZLMQOrkATZLwpgKA&sig2=6Rq7zMpZyNUYi3r0N3MxHQ

Chapter 3
Laser Safety

It seems inevitable that the battlefield laser threat will markedly increase in the coming years. This will be because of not only the development and implementation of laser weapons but also the increasing number of other helpful laser-powered devices such as range finders and target designators. Therefore, it will be necessary for armies to protect their sensors and personnel by introducing passive as well as active countermeasures for laser technology. The primary laser threat will come from laser weapons, although conventional weapons guided to their targets by lasers will also constitute an indirect laser threat, as will be demonstrated later in this chapter.

Protection and countermeasures against laser weapons are difficult problems, which so far have remained unsolved despite years of research. A simple and cheap eye protection against anti-eye laser weapons still does not exist; consequently, protection of personnel involves many complicated factors ranging from filters to defensive battlefield behavior. This chapter will mainly deal with what we can do to protect personnel, sensors, and combat units against the laser beams from low-energy laser (LEL) weapons. Protective measure required to counter high-energy laser (HEL) weapons will only be described briefly.

3.1 Laser Safety

The laser has become a common tool of civilian and soldiers all over the world. Many lasers, perhaps most of them, are in some way dangerous to people. For several reasons, it is our eyesight that is most threatened, but there are many other dangers to deal with as well. Laser safety is very complex problem.

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_3

3.2 Laser Hazards

The use of lasers almost always carries with it some kind of danger, either at the laser site itself or wherever there is a direct, reflected, or scattered laser beam. At the laser site, it is not only the actual laser beam, which can be dangerous, but electrical, chemical, and other hazards exist as well. Most laser power supplies can cause server electrical shocks, possibly even electrocution. Furthermore, many highly explosive and toxic substances are used in solid, liquid, or gas from to power laser cooling systems. For many reasons, it is useful to divide the hazards from laser beam into two main groups: those to the eye and those to the skin. The eyes may be severely damaged and even permanently blinded by rather low-energy laser beams, while the skin is not nearly as sensitive. To get severe skin burns in the visible and infrared part of the spectrum, it is normally necessary to use a very high-energy laser beam, which delivers, at least, several watts per second square centimeter (W/cm^2) to the target. Safety threshold limits both for the skin and eyes are well defined and have resulted in very strict safety regulation.

3.2.1 Laser Hazards to the Eye

To understand laser hazards to the eye fully, with their implications for laser safety requirements and the possibilities of anti-eye laser weapons, it will be discussed in this part briefly. It is necessary to understand the anatomy of the eye, which is beyond the scope of this report, but detail aspect of such discussion can be found in Ref. [1].

The hazardous effects of a laser beam that is transmitted through the eye are, in the vast majority of cases, limited to the retina. The effect upon the retina may range in severity from a temporary reaction without residual pathological changes to permanent blindness. A soldier using magnifying optics may not only be much easier to go blind but may also be a more valuable target than a soldier with a naked eye. Tank gunners, artillery fire controller, forward observer, missile operators, commanders, and others all use magnifying optics in critical moments on the battlefield, and their optical systems may then be detected and identified by the characteristic reflections and exposed to laser radiation. For example, at the microscopic scale, light reflects off locally flat incremental areas of the target skin according to a bidirectional reflectivity distribution function (BRDF) effects, which must be measured and exist inside the materials' database. The detail at the scale of target components, the primitive shapes account for local variations of the surface normal, surface areas, local material assignments, and self-shadowing effects can be found. Within that, the reference author will resolve all geometry effects at this scale through standard ray tracing techniques as part of laser range safety tools (LRST) [2].

The smallest observable reaction may be a whitening of the retina. However, as the retinal irradiance is increased, lesions occur which progress in severity from swelling (edema) to burning (coagulation) and then bleeding (hemorrhage) as well as additional tissue reaction around the lesion. Very high retinal irradiance will cause gas bubbles to form near the site of absorption.

The retina itself is not much more sensitive to laser damage than any other parts of the body. The level of energy that may cause severe damage to any part of the body is between 50 and 500 mJ/cm^2 for a short pulse. It is only the optical concentration of the energy by the optics of the eye that makes a low-powered laser capable of damaging the retina specifically rather than the rest of the eye or body.

The most important part of the eye for vision is the macular area and, in particular, the fovea centralize, which is densely packed with cones. If the laser beam causes a retinal burn of any size in this area, the result is permanent loss of fine-detail vision sufficient to cause legal blindness, and no treatment is possible. Of course, much vision is still present but not enough to read rapidly, drive an automobile, or do any visually demanding task [2].

When discussing laser safety and the eye, it is necessary to differentiate between the effects inside the eye within the retinal hazard region (400–1400 nm) and the effects on the outside of the eye from those laser beams that do not reach the retina [1].

The retinal hazard region covers the spectrum from 400 to 1400 nm and includes the visible and near-infrared parts of the spectrum. The shorter wavelengths in the near ultraviolet are absorbed mainly in the lens, and the even shorter far-ultraviolet wavelengths are absorbed mainly in the cornea. Longer wavelengths in the mid-infrared region are also absorbed in the cornea. Thus, there can be harmful effects to the eye from laser exposures even in the parts of the optical spectrum outside the retinal hazard region [1].

The excessive absorption of intermediate ultraviolet radiation by the cornea causes ultraviolet photokeratitis. This is a very painful but temporary injury, often called snow blindness or welder's flash [1].

However, as most of these effects require that the eye be subjected to high levels of laser radiation for a comparatively long time, half a minute or more, it seems unlikely that these effects in the ultraviolet part of the spectrum could form the basis for a laser weapon.

In the mid- and far-infrared region, the possibility for absorption in the cornea, especially for wavelengths longer than 2000 nm, is very high. Therefore, the cornea is very susceptible to damaging heat during exposure to mid-infrared radiation. If the energy level of the beam is high enough to cause corneal heating, this will produce immediate and severe pain and automatically trigger the blink reflex. The cornea is quite sensitive, and an elevation of only 20 °F in temperature will cause a pain response [2]. The question is whether sufficient thermal energy would be absorbed in the cornea to cause injury in the short time before the blink reflex is activated. The lids are much less sensitive to damage because the circulating blood carries away the heat and a large amount of the laser beam is reflected (Fig. 3.1).

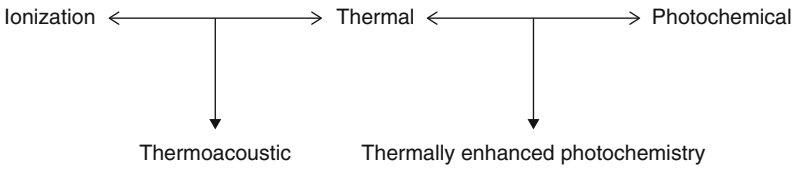


Fig. 3.1 Types of interaction of laser energy with the eye and other biological tissues. Only the thermal and thermoacoustic modes of interaction are important with present-day antipersonnel laser weapons [1]

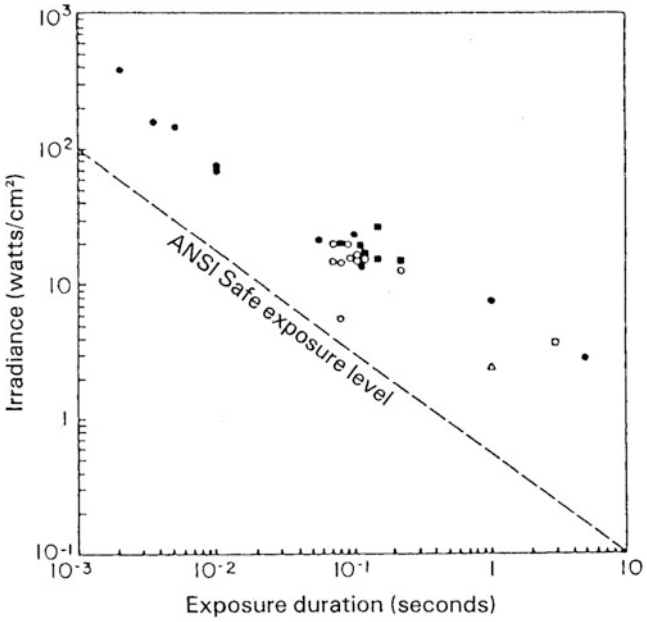


Fig. 3.2 Threshold for corneal injury from CO_2 laser radiation. The differences between data points at the same exposure duration are largely to the use of different corneal image sizes. The data points are from several laboratories and fit a thermal heat flow damage model quite well [1]

The infrared lasers that may be used to injure the cornea are CO_2, hydrogen fluoride (HF), deuterium fluoride (DF), and CO. Such lasers with an output power of more than 10 W/cm^2 could deliver at least 0.5–10 J/cm^2 to the cornea before the blink reflex gives any protection, as shown in Fig. 3.2. Existing infrared lasers can certainly damage the cornea before any head movement can occur. Research has shown that thermal injury to the cornea produces a white spot or an opacification of the surface. The injury is extremely painful and needs immediate and well-qualified medical care. The severity of corneal burn injuries from laser exposure can be compared to that of burns and injuries resulting from ignition or explosion of flammable objects.

3.2.2 Laser Hazards to the Skin

Laser can have several important effects on the skin. The thermal effect is the most significant one. Burn injuries are divided into three basic groups. A first-degree burn is a very superficial reddening of the skin, a second-degree burn produces blistering, and a third-degree burn, the most severe kind, destroys the entire outer layer of the skin. The irradiance necessary to cause a first-degree burn is 12 W/cm^2; for second- and third-degree burns, the necessary irradiation is 24 and 34 W/cm^2, respectively. If the exposure is shortened, the irradiance required to give a third-degree burn is significantly increased. Laser injury threshold for the skin are dependent on the wavelength of the laser as well as on the pigmentation of the skin. Dark skins absorb more and thus get hotter for the same laser energy. For long exposures, the energy levels necessary to produce injury are highly dependent upon exposure duration. It is possible for high-energy lasers to produce significant burns within an exposure period of less than 1 s [1].

A soldier on the battlefield, aware of the threat from laser exposure, will be rather well protected as long as his uniform or the immediate environment is not set on fire. However, it has to be recognized that, even in a protected state, burn injuries to the eyes will probably still be a problem. In far-infrared and ultraviolet regions of the spectrum, where the laser energy does not reach the retina, corneal injury thresholds are approximately the same as for skin injury. Therefore, laser burns to both the exterior of the eye and skin are possible, but these do not seem to be important threats at the moment [1, 2].

3.3 Safety Regulations

There are a variety of laser safety standards including federal and state regulations and nonregulatory standards. The most important and most often quoted is the American National Standards Institute's Z136 series of laser safety standards. These standards are the foundation of laser safety programs in industry, medicine, research, and government. The ANSI Z136 series of laser safety standards are referenced by the Occupational Safety and Health Administration (OSHA) and many US states as the basis of evaluating laser-related occupational safety issues.

ANSI Z136.1 Safe Use of Lasers, the parent document in the Z136 series, provides information on how to classify lasers for safety, laser safety calculations and measurements, laser hazard control measures, and recommendations for Laser Safety Officers and Laser Safety Committees in all types of laser facilities. It is designed to provide the laser user with the information needed to properly develop a comprehensive laser safety program.

For manufacturers of laser products, the standard of principal importance is the regulation of the Center for Devices and Radiological Health (CDRH) and Food and Drug Administration (FDA), which regulates product performance. All laser

products sold in the United States since August 1976 must be certified by the manufacturer as meeting certain product performance (safety) standards, and each laser must bear a label indicating compliance with the standard and denoting the laser hazard classification.

The establishment of the threshold levels for different laser injuries is basic to the whole question of laser safety. The threshold level is an exposure value below which adverse changes have a low probability of occurrence and no significant risk exists. There is always some question about what the actual value of the threshold is, because it varies with the wavelength and exposure duration but also with the individual.

The value of the threshold may be set by using a statistical analysis to determine a certain damage probability (usually 50 %) and then setting the safety level at a selected level of probability below this, usually the 0.01 or 0.001 % level. This energy/power level is often a factor of 10 below that for the 50 % damage point. In order to calculate a correct threshold level, it is also necessary to try to simulate some kind of worst-case scenario—when the eye is hit in its most sensitive part and takes in as much laser light from the laser in question as is possible under the circumstances.

3.4 Laser Hazard Classification

Research studies, along with an understanding of the hazards of sunlight and conventional, man-made light sources, have permitted scientists to establish safe exposure limits for nearly all types of laser radiation. These limits are generally referred to as maximum permissible exposures (MPEs) by laser safety professionals. In many cases, it is unnecessary to make use of MPEs directly. The experience gained in millions of hours of laser use in the laboratory and industry has permitted the development of a system of laser hazard categories or classifications. The manufacturer of lasers and laser products is required to certify that the laser is designated as one of four general classes, or risk categories, and label it accordingly. This allows the use of standardized safety measures to reduce or eliminate accidents depending on the class of the laser or laser system being used. The following is a brief description of the four primary categories of lasers:

Class 1

A Class 1 laser is considered safe based upon current medical knowledge. This class includes all lasers or laser systems which cannot emit levels of optical radiation above the exposure limits for the eye under any exposure conditions inherent in the design of the laser product. There may be a more hazardous laser embedded in the enclosure of a Class 1 product, but no harmful radiation can escape the enclosure.

Class 2

A Class 2 laser or laser system must emit a visible laser beam. Because of its brightness, Class 2 laser light will be too dazzling to stare into for extended periods. Momentary viewing is not considered hazardous since the upper radiant power limit on this type of device is less than the MPE (maximum permissible exposure) for momentary exposure of 0.25 s or less. Intentional extended viewing, however, is considered hazardous.

Class 3

A Class 3 laser or laser system can emit any wavelength, but it cannot produce a diffuse (not mirrorlike) reflection hazard unless focused or viewed for extended periods at close range. It is also not considered a fire hazard or serious skin hazard. Any continuous wave (CW) laser that is not Class 1 or Class 2 is a Class 3 device if its output power is 0.5 W or less. Since the output beam of such a laser is definitely hazardous for intra-beam viewing, control measures center on eliminating this possibility.

Class 4

A Class 4 laser or laser system is any that exceeds the output limits (accessible emission limits, AELs) of a Class 3 device. As would be expected, these lasers may be either a fire or skin hazard or a diffuse reflection hazard. Very stringent control measures are required for a Class 4 laser or laser system.

3.5 Laser Range Safety Tool (LRST) Physics

It may be of interest to consider what the safe distances are for current military laser devices. This will give an indication of the size of the hazardous area around each device. Each specific laser has a safety distance of its own based upon its output properties. The acronym used to describe this distance, in the US regulations at least, is NOHD that stands for nominal ocular hazard distance. The basic definition of NOHD is the axial beam distance from the laser where the exposure or irradiance falls below the applicable exposure limit. The NOHD is calculated to determine at what distance an unprotected person can stand directly in the beam and be exposed momentarily without being injured. The use of magnifying optics must be taken into account, because they will markedly increase the NOHD. A 6-mile NOHD may be increased to 50 miles if an individual looks into the laser with optics that magnify 13 times. It should be remembered that the laser may be hazardous to the eye even beyond the NOHD if the laser is viewed or stared at for a prolonged time. The NOHD is calculated based on momentary viewing only. It is also necessary to take into account the problem of accuracy in aiming the laser. Also, the possibility of reflection is caused by mirrorlike surfaces such as windows, optical surfaces, greenhouse, still ponds, or road signs covered with reflective coating.

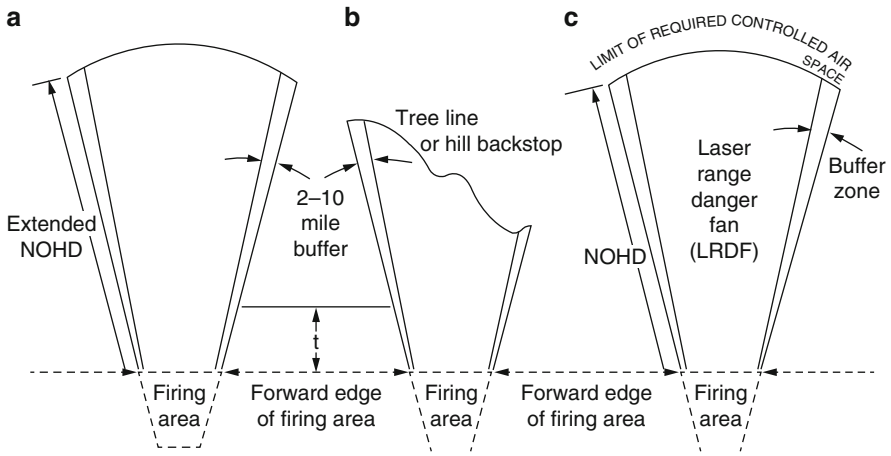


Fig. 3.3 Laser range safety fans. Laser range safety fans are used by the US Army to indicate the nominal ocular hazard distance (NOHD). The NOHD is normally terminated by a backstop. The un-terminated NOHD depends on beam expansion and atmospheric attenuation. In *case A*, the NOHD is a line-of-sight fan parallel to the ground and would only be used when there is no backstop. *Case B* is the more usual situation where a backstop is established by a hill or tree line. In *case C*, the fan is perpendicular to the ground and is applied to airspace hazards [1]

However, for some lasers, there may be a hazardous diffuse reflection area (HDRA), which is typically less than 10 yards from the reflecting surface. An example of a laser danger zone is shown in Fig. 3.3.

Nominal Ocular Hazard Distance and Area

1. **Nominal Ocular Hazard Distance**

Nominal ocular hazard distance (NOHD) is the distance from the source at which the intensity or the energy per surface unit becomes lower than the maximum permissible exposure (MPE) on the cornea and on the skin. The laser beam can thus be considered as dangerous if the operator is closer from the source than the NOHD.

Like the MPE, this distance depends on several parameters:

- The beam characteristics: output power, diameter, and divergence
- The MPE value on the cornea
- Eventually, the optical system inserted in the beam trajectory

For example, this distance can be extremely long for Class 3B and 4 laser sources (see exercises). It is thus necessary to stop the beam at the end of the optical system.

(continued)

When looking at the beam with an optical system, one has to consider the possible higher intensity entering the eye and thus to expand the evaluated NOHD (called afterward expanded NOHD).

As long as the beam propagates freely, this distance can be evaluated according to the following expression:

$$\mathrm{NOHD} = \frac{1}{\theta}\sqrt{\frac{4P_0}{\pi(\mathrm{MPE})} - (2w)^2}$$

In this formula, NOHD is the nominal ocular hazard distance (in meter), P_0 the power of the source (in watts) or eventually the total energy carried by one pulse (in joules), MPE the maximum permissible exposure (in W/rad or $\mathrm{J/m^2}$), w the waist of the Gaussian beam (m), and θ the divergence of the beam.

When using an optical system to look at the beam, one has to take into account the beam focusing induced by the system. Defining f the focal length f of the optical system and α the half-aperture angle of the beam, the expression turns to:

$$\mathrm{NOHD} = f + \frac{1}{\tan\alpha}\sqrt{\frac{P_0}{\pi(\mathrm{MPE})}}$$

2. **Nominal Ocular Hazard Area (NOHA)**

Inside this area, the intensity or the energy per surface unit is higher than the MPE on the cornea. The size of this area is defined by the NOHD. However, it is very difficult to define this area as it depends on the environment (dusty or not) and on the objects than can be on the beam trajectory—in other words, one has to take into account the specular reflections.

The laser beam is decreased or attenuated by some atmospheric conditions, and this is a factor that should be considered when the NOHD is longer than a few kilometers. Atmospheric attenuation is mainly dependent on the sum of three following different effects:

- Large particle scattering
- Molecular scattering
- Absorption by gas molecules (thermal blooming)

Each of these effects is defined briefly in the following paragraphs.

Large particle, or Mie, scattering is the dominant factor in the visible and the near-infrared part of the spectrum, where the particle size of the atmospheric contaminates larger than the wavelength of the laser light.

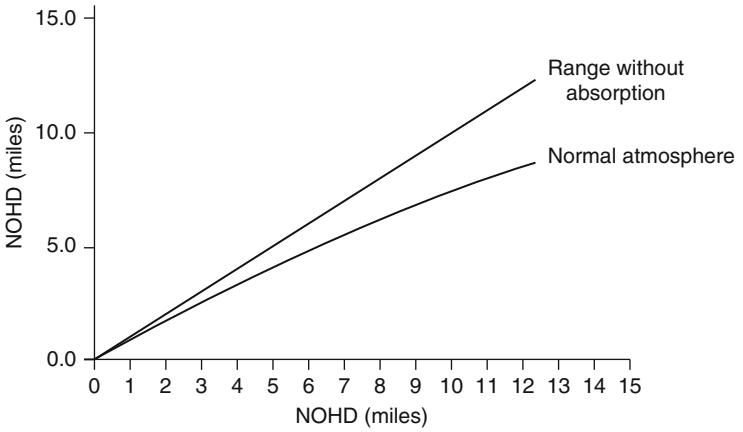


Fig. 3.4 The effects of clear atmosphere on the nominal ocular hazard distance (NOHD). This figure shows the theoretical distance in a vacuum and the actual distance in a clear atmosphere. The *straight line* is for a laser beam at a wavelength (1064 nm) not absorbed heavily by the atmosphere. The *curved line* shows effects of a normal atmosphere. Including that even in the clearest atmosphere, there is a considerable attenuation. As the distances are large, this attenuation will not affect the use of lasers within a typical battlefield but will modify the effects of lasers from aircraft or in an antiaircraft situation [1]

Molecular or Rayleigh scattering by oxygen, nitrogen, and other molecular constituent and, in these cases, the molecular size are much less than wavelength of laser beams.

The contribution of absorption by gas molecules and other particles to attenuation is most important in the infrared region of the spectrum.

The molecular scattering of the laser beam increases at shorter wavelengths. However, this effect is not substantial over short distances. A normal and clean atmosphere is relatively transparent to the argon laser beam (blue), the ruby laser beam (red), and the Nd:YAG beam (near infrared). If NOHD calculated for vacuum transmission is known and compared to the NOHD compensated for the ambient atmosphere (see Fig. 3.4), it may be concluded that, for low-energy lasers, the atmosphere even at battlefield distance up to 10 miles, at least, is not a big problem. A ruby laser beam might be attenuated as little as 10 % at 6 miles [2].

High-energy laser (HEL) beams are much more heavily dependent on the weather conditions, especially rain, snow, dust, clouds, and smoke. This will be discussed further in Sect. 3.4. However, what may be concluded from laser safety calculations is that most military concluded from laser safety calculations is that most military lasers are not very dependent on the atmosphere when the air is fairly clean and the laser is used at battlefield distances.

Therefore, LRST of target reflectivity known as BRDF [3] (bidirectional reflectivity distribution function) database should be reviewed and updated for the target that laser weapon is engaging. Continuous testing for LRST should be conducted,

and desired results for each test condition must be studied considering the following criteria:

- LRST vs. hand calculation radiometry
- Check of ANSI maximum permissible exposure (MPE)
- LRST hazard zone vs. hand calculation hazard zone

The American National Standards Institute (ANSI) initiated the first work on a comprehensive standard for the safe use of lasers. Initial hypothetical scenario for LRST should be assumed, and test results should be compared with those which are hand calculated using various related computer codes, and we hope to be at least within optimum design range or at least well within uncertainties built into ANSI standard.

A safety analysis of the outdoor use of current military range finders and target designators against the eye of the human being such as soldiers. However, so far as it is publicly known, only one country has made such a use as part of its military practice. This is the case of the use by the British of low-energy lasers to flash blind Argentinean pilots in the Falklands conflict. At present, there are no known unclassified tactical manuals which cover the deliberate use of lasers against the eyes of soldiers as a weapon of warfare or reflection of a beam in term of BRDF from a target engagement. What the military will do in future conflicts, which will certainly involve a mass use of lasers, remain to be seen.

The underlying requirements of the laser range safety tool (LRST) and its written code will be to generate an accurate time history of intensity at potential positions where humans might be exposed to reflected laser beam light from a target missile being irradiated in the case of airborne laser (ABL) [2].

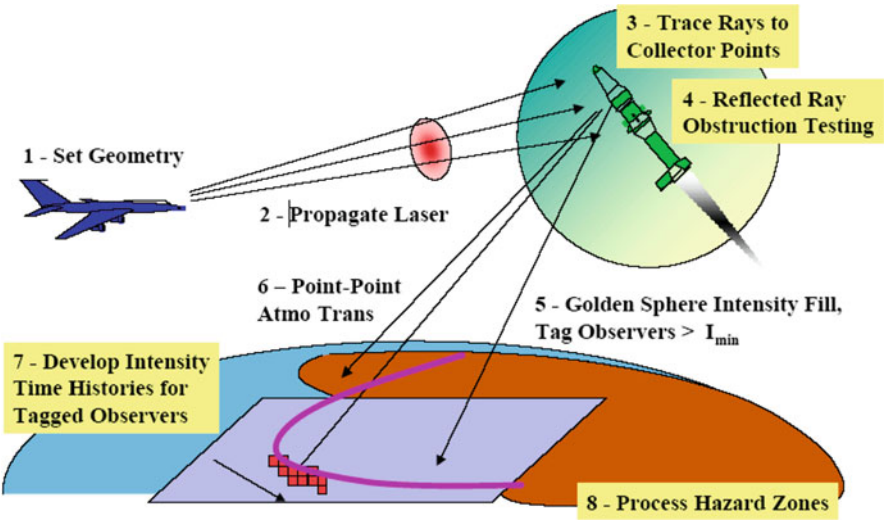

3 - Trace Rays to Collector Points
4 - Reflected Ray Obstruction Testing
1 - Set Geometry
2 - Propagate Laser
6 – Point-Point Atmo Trans
5 - Golden Sphere Intensity Fill, Tag Observers > I_{min}
7 - Develop Intensity Time Histories for Tagged Observers
8 - Process Hazard Zones

Fig. 3.5 Eight basic algorithmic steps to evaluate hazards

There are about eight basic algorithmic steps necessary to evaluate eye hazards in LRST. Figure 3.5 shows a simple diagram of this process.

Certain conditions, preengagement, and dual on target time of laser beam scenario and its geometry with target in battlefield must be calculated. In such thing as the position and velocity of ABL, an incoming threat also should come under consideration as well as environmental condition has to be part of the parameter of engagement equation.

The incoming threat such as in flight missile position and velocity should be derived from database table of pre-simulated position is imported into code as part of solution to LRST consideration. All the vector position of missile in relative to itself and ABL that is known and referred to as golden sphere has been considered in the LRST code simulations [2].

Overall the LRST is a software tool for calculating reflected energy hazards under the role of BRDF in the context of test range development of the airborne laser (ABL) system. The LRST software package consists of a suite of programs that provide facilities for describing and simulating test scenarios, including the range, the ABL, the aircraft laser complement, the target, and the observers. What is missing from this software is the effects of BRDF on friendly troops on the ground and preengagement warning that ABL will be in tangle with the incoming missile threat. The other missing link of the LRST is the consideration of falling debris due to physical destructive ABL impact on target and dual time on target. But the LRST software does an excellent job of providing a detailed description of physical basis of the simulation and the computational algorithms utilized to calculate the radiometry and hazards.

References

1. Sliney D, Wolbarsht M (1980) Safety with lasers and other optical sources: a comprehensive handbook, 1st edn. Springer, New York
2. Crokett G (2003) Laser range safety tools (LRST) physics reference. Logicon-RDA, Albuquerque, NM. Report AFRL-HE-BR-TR-2003, September 2003
3. Crokett G (1999) Bi-directional reflectivity distribution function (BRDF) modeling to LRST: Maxwell-Beard, Phong, and Gaussian models. Contract F04701-98-D-0100, CDRL A004, January 1999

Chapter 4
Laser Weapons

Laser technology is only 30 years old, but it is much diversified. There are already varieties of military applications, although there are many limitations restricting the use of lasers. Today, the armed forces in most countries routinely use a wide range of laser devices such as laser range finders and designators. In some countries, work is proceeding on more imaginative laser weapon concepts that will eventually fulfill realistic, yet very precise, military requirements. The design of a specific laser weapon is heavily influenced by the characteristics of the intended target. If the desired effect of the weapon is to neutralize aircraft, helicopters, or missiles by burning holes through them or tanks by putting many miniature cracks (crazing) in the glass vision blocks to make them appear to be frosted, a very high-energy laser has to be used with a power output on the order of several megawatts (MW). Such a laser would be a true anti-material weapon. However, if the target is a sensitive electro-optical system or some other type of sensor system, which has to be jammed or destroyed by a laser operating in a countermeasure mode, the choice will be a low-energy laser operating within the frequency bandwidth of the target sensor. This use of a laser can also be considered anti-material. If the target is a soldier, there is one part of his body that is extremely sensitive to laser radiation—his eyes. It is sufficient to use a low-energy laser operating in the visible or near-infrared (near-IR) part of the spectrum to damage the soldier's eyes and, in effect, cause blindness. If the laser is to cause burn injuries to the soldier's skin or to set fire to his uniform, a high-energy laser is required. In either case, if the purpose of the laser is to blind or burn the soldier, it will obviously be antipersonnel.

4.1 Laser as a Weapon

Even before the laser was invented, science fiction writers told of incredible weapons and machines that emitted a bright saber of light, a death ray that disintegrated everything in its path. Even today, science fiction movies and books

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_4

place high emphasis on weapons that use light instead of bullets. The laser beam is popularly thought of as a very powerful death ray which can be fired from a handheld laser gun to vaporize soldiers, demolish building, and burn-through target armors. In reality, the laser is a suitable tool for many military applications and can be turned into a deadly weapon, but there are definitely limitations to what a laser can do. The laser really is a ray weapon, and its light rays can damage some targets in a way that appeals to the most vivid imagination. It is important to take these somewhat speculative factors into consideration when studying the psychological effects of the use of laser weapons on the battlefield. Otherwise, it will not be possible to get a complete and realistic picture of what using a laser really means to the combatants.

4.2 Possible Targets

A discussion of laser weapon applications outlining what laser weapons can really do must start with the destination of the laser beam—the target. The desired effect on the target ultimately decides what is needed from the laser. To a large extent, the interaction between the laser beam that is selected and the target also determines which cost-effective weapons are developed, produced, and deployed into the battlefield.

The sensitivity of the target to laser light determines whether a low-energy or high-energy laser is required. If the target is sufficiently sensitive to low levels of energy within a comparatively broad band of the spectrum, a cheap and cost-effective laser weapon can be designed and mass-produced. If high energy is required, the possibility of designing a usable and affordable laser weapon decreases drastically.

4.3 Energy Level at the Target

One of the basic questions facing the laser weapon designer is what energy level must be absorbed by the target in order to get the desired result. The absorbed energy (E) is some fraction (A) of the product of the power density or intensity (I) present in the laser beam and the emission duration (t). E is measured in energy units, joules (J) or watt seconds per area, usually expressed in square centimeters; I in power units, watts (W) per square centimeter; and the time in seconds in the following equation:

$$E = A(I \times t)$$

This means that if the emission duration is required to be short, as it would be in the engagement of multiple targets, the power density has to be as high as possible. The

power density is calculated as the beam power divided by the size of the "beamed" area, which means that a high beam power and a small surface area will give a high power density. How much of the laser power will finally be absorbed by the target in the affected surface area will determine what destructive effect will be achieved. The laser power goes from the laser to the target, suffers transmission losses in the optical system and the atmosphere, and has a further loss when some of the power is reflected from the target surface. The absorbed power is normally no more than 20–60 % of the original emitted laser power.

One parameter that is useful in determining the effectiveness of a Gaussian laser beam is the beam irradiance at the target. For a beam with output power P_0 and cross-sectional A at the target, the peak irradiance I_p at the target is

$$I_p = \frac{P_0 \tau}{A}$$

where τ is the atmospheric transmittance.

The effectiveness of a laser beam in causing mechanical damage is, thus, dependent on beam power, pulse duration, wavelength, air pressure, the material, and the finish of the target surface. For example, a painted area has considerably increased energy absorption when compared to an unpainted aluminum plate. The absorption varies widely between different materials and at different wavelengths. The absorption of a ruby laser at 694 nm is 11 % for aluminum, 35 % for light-colored human skin, and 20 % for white paint. The corresponding figures for a CO_2 laser at 10,600 nm are 1.9, 95, and 90 %. This also indicates that one way to counter a HEL weapon is to choose a very reflective material for the target surface. On the other hand, longer wavelengths emitted by the laser can reduce the effects of highly reflective materials and increase the absorption. Every factor in this very difficult pattern combines to determine the degree of target destruction as well as the final energy level that will be needed to produce the desired effect.

It is obvious that the level of energy required to destroy a target varies considerably depending on the circumstances. Therefore, it is not surprising that the required energy level figures quoted in the open literature also show rather large variations. In spite of this, some numbers may be given which indicate the general range of energy levels.

An aircraft, helicopter, or missile could be hit with an HEL weapon in many different ways that in the end would nullify it. Fuel tanks could be ruptured, or the fuel itself could be caused to explode. Windshields could be shattered, and parts of the control surfaces such as elevators or rudders could be destroyed or disturbed enough to make it impossible to continue fighting. The rotor head of a helicopter or the wing of an airplane or missile could be made to fail, resulting in a crash. Sensors, radars, and other navigation aids could be destroyed; if this destruction occurs during a sensitive and crucial moment in the last phase of an attack, it could result in a crash or an aborted mission. Also, in some situations, an HEL weapon could even explode the ammunition carried by an airborne attacker.

To punch through the metal skin of an airplane requires about 700 J/cm^2, although it should be noted that a hole burned in the skin of an airplane may not be sufficient to destroy it in the air or even to make it crash. A more realistic energy level to disable an aircraft may be five to ten times higher, which means that a successful HEL weapon will have to be able to deliver at least 5000–10,000 J/cm^2 on the target.

Optical sensors and radomes (plastic radar domes) are much easier to damage; no more than 10 J/cm^2 needs to be delivered directly on the target. Furthermore, if the laser wavelength is within the sensitive wavelength region of the sensor in question, the energy needed could be extremely low. If the HEL weapon is used as an antipersonnel weapon, that is, as a long-range flamethrower, the energy necessary to burn exposed skin is merely 15 J/cm^2, and damage to the cornea, the clear window into the eye, requires only 1 J/cm^2.

4.4 Absorption and Scattering

The Earth's atmosphere is acting like an absorbing medium. Absorption occurs when a photon of radiation is absorbed by a gaseous molecule of the atmosphere that converts the photon into the molecule's kinetic energy. Hence, absorption is a way and procedure by which the atmosphere is heated, and it is a strong function of laser or radiation of wavelength. For example, propagation of radiation essentially gets eliminated at a wavelength below 0.2 μm due to absorption by O_2 and O_3, while there is very little absorption at the visible wavelengths (0.4–0.7 μm).

Scattering of electromagnetic waves in the visible and IR wavelengths occurs when the radiation propagates through certain air molecules and particles. Light scattering is strongly wavelength dependent, but there is no loss of energy like in absorption. The physical size of the scatters determines the type of scattering:

- *Rayleigh Scattering*—This is named after Lord Rayleigh and caused by air molecules and haze that are small in comparison with the wavelength λ of the radiation (see Fig. 4.1 below). Rayleigh scattering, also called *molecular scattering*, applies only to very clear atmosphere. The scattering coefficient is proportional to λ^{-4}, a relation known as the *Rayleigh law*. For these small air molecules, scattering is eligible at wavelengths greater than roughly 3 μm. At wavelengths below 1 μm, Rayleigh scattering produces the blue color of the sky

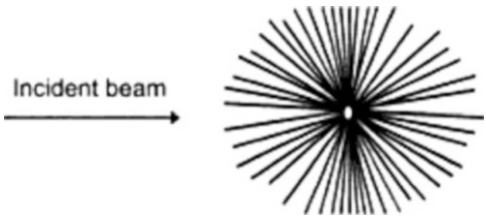


Fig. 4.1 Rayleigh scattering

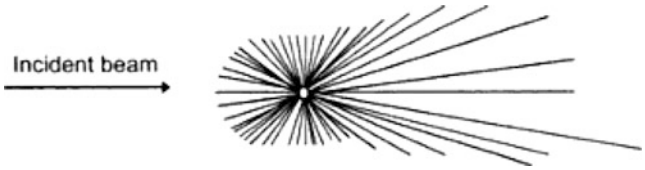


Fig. 4.2 Mie scattering

as a consequence that blue light is scattered much more than other visible wavelengths.

- *Mie Scattering*—(named after Gustav Mie) scattering by particles comparable in size to the radiation wavelength (also called *aerosol scattering*). Unlike Rayleigh scattering, scattering by particles comparable in size to or greater than the radiation wavelength is concentrated in the forward direction (see Fig. 4.2). Scattering losses decrease rapidly with increasing wavelength, eventually approaching the Rayleigh scattering case. Mie scattering is the reason why sunset appears red.

A term that is sometimes used to describe atmospheric visibility is the visual range, which corresponds to the range at which radiation at 0.55 μm is attenuated to 0.02 time that is a transmitted level. Rayleigh scattering by molecules implies a visual range of approximately 340 km (or 213 miles) [2].

Absorption and scattering are often grouped together under the topic of *extinction*, defined as the reduction or attenuation in the amount of radiation passing through the atmosphere. The *transmittance* (also called *atmospheric transmission*) of laser radiation that has propagated a distance L is related to extinction as described by Beer's law, which can be written as [1, 2]

$$\tau = \exp[-\alpha(\lambda)L] \quad \text{unitless}$$

where $\alpha(\lambda)$ is the extinction coefficient and the product $\alpha(\lambda)L$ is called the *optical depth*. The extinction coefficient is composed of two parts:

$$\alpha(\lambda) = A_a + S_a \quad [\text{m}^{-1}]$$

where A_a is the absorption coefficient and S_a is the scattering coefficient. Absorption and scattering are deterministic effects that are fairly well known.

Software packages like LOWTRAN, FASCODE, MODTRAN, HITRAN, and PcLnWin (most of these codes are available from Galaxy Advanced Engineering) are commonly used by both government and private industry to predict transmittance (attenuation) effects as a function of wavelength λ, based on a variety of conditions—meteorological range, latitude (tropical, mid, arctic), altitude, etc. A typical output from MODTRAN for rural aerosols with meteorological range of 23 km is shown in Fig. 4.3 as a function of wavelength over 1–10 μm.

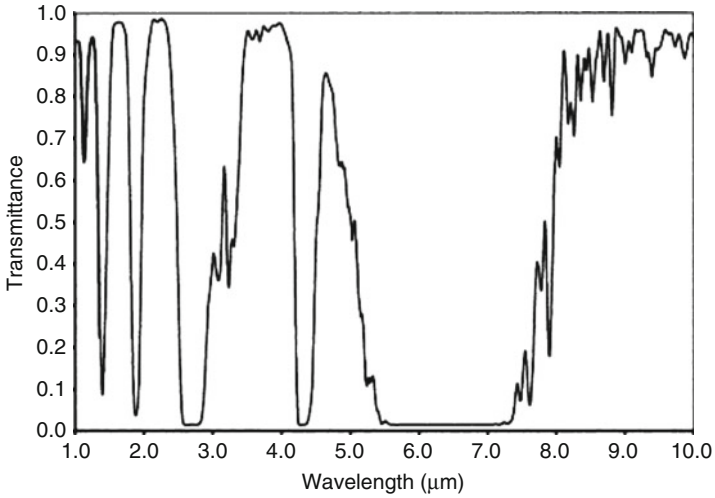


Fig. 4.3 Typical atmospheric transmittance for a horizontal 1-km path. The height above ground is 3 m with no rain or clouds

4.5 Atmospheric Structure with Altitude

The atmosphere is a gaseous envelope that surrounds the Earth and extends to several hundred kilometers above the surface. Over 98 % of the atmosphere by volume is comprised of the elements nitrogen and oxygen. The major constituents of the atmosphere are water vapor, carbon dioxide, nitrous oxide, carbon monoxide, and ozone. Based mostly on temperature variations, the Earth's atmosphere is divided into four primary layers (Fig. 4.4):

- *Troposphere*—extends up to 11 km and contains roughly 75 % of the Earth's atmospheric mass. Maximum air temperature occurs near the surface of the Earth, but decreases with altitude to −55 °C. The *tropopause* is an isothermal layer extending 9 km above the troposphere where air temperature remains constant at −55 °C. The tropopause and troposphere together are known as the *lower atmosphere*.
- *Stratosphere*—the layer above the tropopause, which extends from 20-km up to 48-km altitude. The air temperature is roughly constant in the very lowest part of the stratosphere but then increases with altitude because the ozone gas in this layer absorbs ultraviolet sunlight, thereby creating heat energy. The ozone layer, which protects life from harmful ultraviolet radiation, is concentrated between 10 and 50 km. Separating the stratosphere from the mesosphere is the stratopause, another isothermal layer at approximately −3 °C.
- *Mesosphere*—extends from the stratopause to roughly 80 km. The temperature here generally decreases at a constant rate down to −90 °C, which is the coldest temperature in the atmosphere. The *mesopause* is the third isothermal layer,

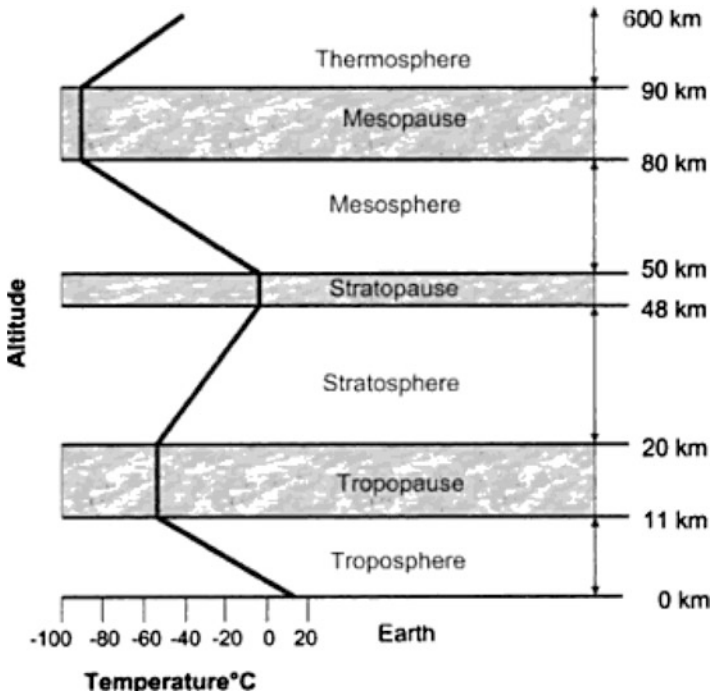


Fig. 4.4 Diagram depicting various atmospheric layers and air temperature

separating the mesosphere, along with the stratopause and mesopause, constituting what is commonly called the *middle atmosphere*.

- *Thermosphere*—extends from the mesopause to roughly 600 km. Air temperature in the thermosphere increases quite strongly above 90 km due to the sun's energy. Most of the *ionosphere* and the *exosphere* are included in the thermosphere. The ionosphere starts around 70 or 80 km up to an indefinite height (~1000 km) and is so named because it is sufficiently ionized by solar ultraviolet radiation that the concentration of free electrons in this layer affects the propagation of radio waves.

4.6 The Major Laser Weapon Concepts

There is generally more than one laser weapon alternative for each proposed laser weapon mission on the battlefield. It is quite possible to vary the laser properties and energy level, the tracking system, and the fire control equipment according to the military requirements for each specific mission. Environmental influences will also have a very strong impact on the final choice of laser weapon applications. For example, hydrogen fluoride (HF) laser is not the best choice for long-range use

within the atmosphere, because its wavelength is strongly absorbed by the atmosphere. Every laser weapon that is designed to operate within the atmosphere over any great range, whether ground-based laser (GBL), sea-based laser (SBL), or air-based laser (ABL), must use wavelengths at which the atmosphere absorption and scattering are as small as possible [2].

To be effective, the wavelength of a laser weapon must be short, at least in the visible band, but preferably in the ultraviolet or X-ray band. The greatest difficulty in designing short-wavelength lasers is power—the shorter the wavelength, the more energy that is required. Optical (visible or ultraviolet) lasers work by heating the skin of the target. The beam must remain at the same spot for several seconds until the skin is hot enough to do internal damage to the target. This is tough because the typical ballistic missile travels in excess of 6 miles per second. Imagine focusing on the same 2′ or 3′ spot over a distance of 50,000 ft, and you have an idea how accurate such a laser weapon must be.

In addition to the problems of accuracy, laser weapons of any power tend to be monstrous, and there are many technical obstacles that the designer should overcome. The SBL is using relay mirrors to direct the beam to the target. ABLs are using turbine-powered chemical jets, and they are placed aboard of aircraft, but the wavelength of the light is long—6–10 μm—far in the infrared region. This makes laser relatively inefficient at destroying their targets unless certain atmospheric and environmental condition met for target engagements.

X-ray lasers, still wrapped in secrecy, emit an extremely high-powered beam that can literally destroy a missile in mid-flight. X-rays can't be deflected by mirrors, however, which means that the weapon must be easily aimed and in a direct line of sight to the target. Fortunately, X-ray lasers can be built small, experts say, making them suitable for space-based operation. The biggest disadvantage to X-ray lasers is that they use an internal nuclear explosion to work, so they are essentially one short device.

A relative newcomer to this laser weapon scene is the free-electron laser, which is being developed at the several national laboratories and universities. The free-electron laser (FEL) uses a stream of electrons that is made to emit photons of light after being oscillated by giant electromagnetic. Free-electron lasers (FELs) have been built and they do work. However, if put into production, an actual antiballistic missile FEL would take up a huge field such as football field or more to function. Obviously, such a device would be useful only as a stationary ground-based laser (GBL) weapon with its present technology.

Laser weapons may be used within an army's air defense against aircraft, helicopters, and missiles. The desired effect on the target may either be to burn holes or destroy key structures, to blind or trick the sensors, or to blind the crews temporally or permanently. The high-energy air defense laser may use all three effects at the same time if the target is within reach of the main effects of the laser. At longer distances, only the anti-sensor and anti-eye capability will be possible. The low-energy air defense laser will use enough energy to be effective against sensors and eyes. It is also possible to field a laser with the main purpose of blinding

or flash blinding the crews. Flash blinding will be most effective in the dark when the eye is dark adapted and much more sensitive to overload by bright flashes [2].

4.7 Small-Scale Weapons Using Lab-Type Lasers

So far part of this report has covered high-energy weapons, designed to counter a major military conflicts and attack. Laser guns in the movies are often handheld devices or, at most, small enough to prop on a vehicle. Lasers powerful enough to inflict damage, but small enough to be carried, have developed, but they are not used in any current military application. It's relatively easy, for example, to build a handheld ruby laser that puts out bursts of large amounts of light energy. When focused to a point, the light from a ruby laser can cut through the paper, cloth, skin, or even thin metal.

Ruby crystals are poor conductors of heat, so ruby lasers emit only short pulses of light to allow the crystal to cool between firings. Nd:YAG lasers operate in a similar fashion as ruby lasers, but they can produce a continuous beam. Making a handheld Nd:YAG laser is no easy feat either; however the Nd:YAG crystal must be optically pumped by another high-powered laser or by an extremely bright flash lamp or light source. Though the power output of an Nd:YAG laser is extremely high, considering the current state of the art, a handheld model is impractical. However, such a weapon could be built as a "laser canon," transported on an armored vehicle or on a towed trailer.

CO_2 lasers are often used in industry as cutting tools. This type of laser is known for its efficiency—30 % or more compared to the 1–2 % of most gas and crystal lasers. A pistol-sized CO_2 laser would probably be difficult to design and manufacture because the CO_2 gas mixture (which includes helium and nitrogen) must be constantly circulated through the tube. What's more, the laser requires a hefty electrical power supply. Still, such apron could be built in an enclosure about the same size as a personal rocket launcher. These are designed to be slung over a shoulder and fired when standing in an upright position [3].

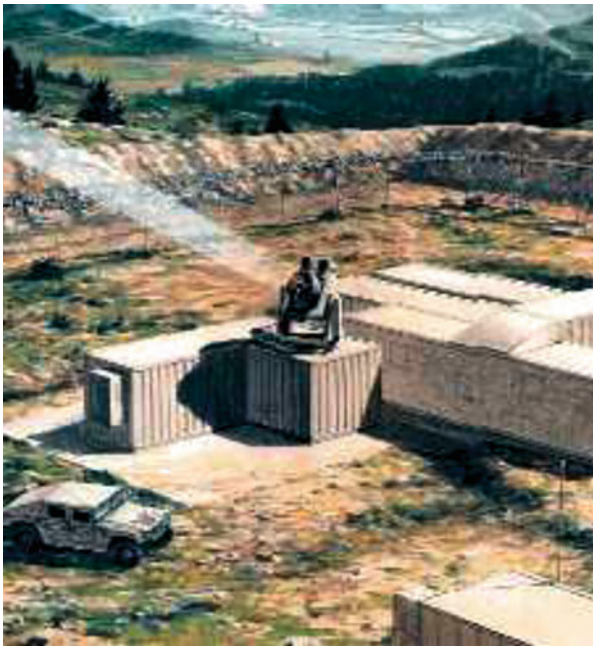
4.8 High-Energy Lasers as Weapons

An air defense HEL weapon designed to shoot down airplanes, helicopters, and missiles successfully must have the ability to keep a very powerful beam at one point on the target for a long enough time to deliver at least 5000 J/cm^2. This requires a laser in the megawatt range. If the shot is to be successful, it must be directed to a certain part of the target that is limited in size and very sensitive and then kept there until the desired effect is reached. Thus, the laser beam must track and follow a target if any great length of time is needed to achieve the desired effect.

Many parts of an aircraft or helicopter are highly resistant to an HEL weapon, but there are still enough thin-skin parts and sensitive areas to produce a devastating effect or destruction if hit precisely. On the other hand, it is obvious that at battlefield ranges even an extremely high-energy laser weapon cannot penetrate the heavy armor on a tank or other armored vehicles and thus an HEL weapon is of no use for destroying resistant ground targets in the battlefield. However, sensors, optics, and related devices are still valid targets wherever they appear on the battlefield, even in a tank.

4.9 High-Energy Laser (HEL) Safety Program

As high-energy lasers move from the safe confines of the laboratory into the outdoor domain, new problems arise in dealing with laser safety. The Tri-Service Laser Bioeffects program at Brooks AFB will be examining the safety aspects of the new technologies and weapon systems to be employed in the future. Several new high-energy laser systems are now scheduled for deployment in the immediate future.



The Tactical High-Energy Laser (THEL) will be a ground mobile system which will use a chemical laser to destroy low-flying threats. Currently under development in a cooperative program with Israel, the THEL conducted test firing in FY1998 and continues several test phases at White Sands Missile Range today.

The airborne laser (ABL), with its megawatt-class laser systems, will engage tactical ballistic missiles during boost phase at altitudes over 40,000 ft. Its lasers will be test fired in FY2003, with Engineering and Manufacturing Development scheduled to begin soon after that. The high-power chemical oxygen iodine laser will have nominal eye-safe distances in the order of thousands of miles and will produce reflection patterns off targets that also have the potential for causing eye damage at very large distances.

Combining these laser characteristics with the planned usage of these weapon systems involving moving targets and possibly moving sources results in a complex series of laser safety calculations to allow safe testing and usage of high-energy lasers outdoors. The Air Force Research Laboratory, Optical Radiation Branch, is developing new tools and techniques for calculating laser hazard areas which include computer modeling of the interaction of high-energy lasers and moving targets and the use of probabilistic methods to augment deterministic calculations.

One of the Missile Defense Agency's highest priority programs involves putting a weapon-class laser aboard a modified Boeing 747-400 series freighter aircraft and using that laser to destroy ballistic missiles shortly after launch. The program is called the airborne laser, and its development could forever change the way that nations wage war.

4.9.1 Airborne Laser (YAL-1A)

Destroying ballistic missiles is a complicated process, one that is confounded even more by the revolutionary use of a directed energy device as a weapon rather than as a targeting or range-finding apparatus. To be successful, ABL must:

- Be housed aboard a stable platform that can stay aloft for hours on the end above weather systems whose clouds could refract its laser beams and nullify its effectiveness.
- Be equipped with sensors able to locate a ballistic missile shortly after launch and hold the track long enough for other system elements to swing into operation.
- Be implemented with a sophisticated computer system capable of keeping track of dozens of missiles and prioritizing them so that the most threatening is targeted first.
- Have a highly developed optical system capable of measuring the amount of thermal disturbance between the aircraft and the target; then be capable of directing a beam of energy that self-compensates for the clear-air obstacles.
- Possess the ability to focus the killer beam on a rapidly rising target, which may be traveling at a speed of Mach 6 or more; then keep the shaft of energy in place long enough to burn a hole in the missile's metal skin.
- And lastly, be provided with a laser powerful enough to prove lethal at a distance of hundreds of kilometers.

Some of those requirements already have been tested:

- The first ABL aircraft—YAL-1A—made its virgin flight over the western Kansas on July 18, 2002, staying aloft for 1 h and 22 min before returning to the Boeing modification facility in Wichita. Between then and the time it transitioned to its new temporary home at Edwards Air Force Base, California, in December, YAL-1A made an additional 13 flights logging more than 60 flight hours.
- As part of a Missile Defense Agency test over the Pacific Ocean in December 2002. ABL's infrared trackers successfully detected a Minuteman booster rocket as soon as it broke the clouds, holding a lock until the rocket's engines burned out 500 km downrange.
- Its battle management (computer) system was flight tested in late summer and early fall of 2002 to verify internal crew communications and the V/UHF radios, plus the data acquisition system and high-definition VHS.

- The six infrared search and track sensors were successfully flight tested.
- The first COIL module that will be installed on YAL-1A tested at 118 % of anticipated power during a shakedown run at TRW's facility in San Juan Capistrano, Calif., in January 2002. Shortly afterward, it was disassembled and shipped to Edwards Air Force Base.

In December 2002, YAL-1A was pulled into a hangar at Edwards' Birk Flight Test Facility where it will be grounded, while the lasers and optical components can be tested and installed.

The goal of the Missile Defense Agency, which has overall management responsibility for the program, and the Airborne Laser System Program Office at Kirtland Air Force Base, N.M., is to have YAL-1A ready to shoot down a threat-representative ballistic missile by December 2004. Currently, the missile is scheduled to be launched from Vandenberg Air Force Base, Calif., with the shootdown to take place over the Pacific.

Construction and testing of YAL-1A (prototype attack laser, model 1A), the first aircraft in a proposed fleet of so-far undetermined size, are the results of an effort by MDA, the program office, the Air Force, and three major contractors—Boeing, Lockheed Martin, and Northrop Grumman Space Technologies (formerly TRW). In addition, the US Air Force's Aeronautical Systems Center, headquartered at Wright-Patterson Air Force Base, Ohio, has provided office personnel. The Air Combat Command, headquartered at Langley Air Force Base, Va., will assume control over the plane once it is declared operational and transferred back to the Air Force.

The ABL program office was formed in 1993. Three years later, in November 1996, the Air Force awarded a $1.1 billion contract to the Boeing Defense Group of Seattle, Wash.; TRW Space and Electronics Group of Redondo Beach, Calif.; and Lockheed Martin Missiles & Space of Sunnyvale, California.

Boeing built the aircraft in Everett, Wash., and modified it in Wichita. The company also developed the hardware/software used in the battle management system and is managing the integration of the main components. TRW built the megawatt-class COIL laser that produces the knockout punch to ballistic missiles, and Lockheed Martin is responsible for the optical system.

Another key organization is the Air Force Research Laboratory's Directed Energy Directorate, also at Kirtland Air Force Base, N.M., where the COIL was invented in 1977. For a quarter of a century, the Laboratory has been conducting research into a myriad of technologies needed to make a laser-carrying aircraft a reality. Besides the COIL, the Laboratory also developed the technologies that will increase the distance laser light can travel through the atmosphere to destroy attacking missiles.

The Aircraft—The Air Force bought a 747-400F straight off the Boeing Commercial Aircraft assembly line and flew it to Wichita, Kan., in January 2000. Boeing workers virtually rebuilt the aircraft, installing miles of wiring, grafting huge sheets of titanium to the plane's underbelly to protect the exterior from the heat of the laser exhaust, and, most importantly, adding a 12,000-lb bulbous turret on the front of the aircraft to house the 1.5-m telescope through which the laser beams will be fired. Company officials said it was the largest military modification to a commercial aircraft that Boeing had ever attempted.

Acquisition, Tracking, and Pointing—In addition to a powerful laser, an airborne laser system also must be able to find and hit its targets. Numerous tests have been conducted at the White Sands Missile Range in southern New Mexico, both with lab-type instruments and with the actual aircraft, to demonstrate the system's ability to identify and follow a potential target.

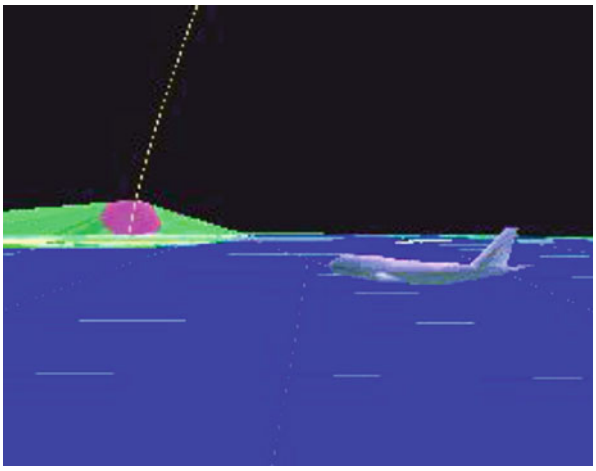
The Lasers—Central to this system is the COIL. As a laser that generates its energy through chemical reaction, it has advantages over solid-state lasers, most notably in the amount of energy it can produce. COIL energy is produced by chemical reaction when oxygen and iodine molecules are mixed. A tremendous additional advantage is that the laser propagates at 1.315 μm in the infrared (invisible) spectrum. This wavelength travels easily through the atmosphere and has greater brightness—or destructive potential—on the target. There are three other important lasers aboard the aircraft: the Active Ranger System, which provides preliminary tracking data; the Track Illuminator Laser, which produces more refined data; and the Beacon Illuminator Laser, which measures the amount of atmospheric disturbance.

Correcting For Atmospheric Turbulence—The ability to find and track a boosting missile would be meaningless without a corresponding ability to lock onto and destroy the intended target. Since air, like water, is made up of many layers, scientists needed to find a way to compensate for these disturbances in the atmosphere in order to focus a high-energy beam on the target and hold the beam in place long enough for it to complete the destruction process. The system that will be installed on YAL-1A is the result of more than 15 years of research conducted by scientists at the Laboratory's Directed Energy Directorate and the Massachusetts Institute of Technology's Lincoln Laboratory. Working out of astronomical facilities at the Starfire Optical Range in the southeastern corner of Kirtland Air Force Base, researchers made revolutionary breakthroughs using lasers, computers, and deformable optics.

The ABL Integrated Test Force—Actually a complex of buildings located at the historic Birk Flight Test Facility at Edwards Air Force Base, Calif. The gem of the ITF is the System Integration Laboratory (SIL), an 18,000-square-foot building housing a surplus 747-fuselage test stand that will serve as a laser template for the ABL aircraft. The six modules that compose the COIL component initially will be tested in the SIL. Once those tests have been completed, the modules will be disassembled and then reassembled on YAL-1A. Other resources in the ITF complex include a ground-pressure recovery assembly (GPRA), which will enable simulation of ABL's anticipated cruising height, and a mixing area for basic hydrogen peroxide, a vital ingredient to the main laser's chemical reaction process.

History—Almost 20 years ago, the Air Force Research Laboratory and its predecessor units completed a project that showed the potential for an airborne laser. A KC-135A tanker airplane (a military version of the Boeing 707) was modified and equipped with a gas dynamic laser. This aircraft shot down a low-flying drone and five air-to-air missiles, proving the concept was possible. Later tests also were conducted at White Sands Missile Range aimed at finding out how effective a laser would be. For these tests, the nation's most powerful laser, the Mid-Infrared Advanced Chemical Laser, was used. In every case, scale models of typical targets were easily destroyed.

The System—Computer simulations indicate that an airborne laser would be very effective under battle conditions. Currently, the program will provide the United States with its only near-term boost-phase missile defense, that is, the ability to find and destroy a missile between the time it is launched and its booster rockets burn out.



The laser range safety tool (LRST) is being developed to permit range safety officers to properly assess hazards and configure test scenarios such that these new weapon systems can be tested in a safe manner. For any given scenario, the

tool assists in the evaluation of reflection patterns resulting from targeting various types of moving targets and predicts hazard zones and appropriate "keep-out" areas. The tool is based on the ANSI Z136.1 standard for eye safety calculations along with new bioeffect data for the specific wavelengths associated with the new high-powered laser systems. Moreover, new risk assessment techniques using probabilities associated with the various aspects of exposure and injury will provide more realistic predictions for use by the range safety officers, as well as the operational users.

4.9.2 Tactical High-Energy Laser for Air Defense

The US Army Space and Strategic Defense Command (SSDC) is working on a new active defense weapon system concept to combat the threat to our forces from so-called dumb munitions. The command's mobile Tactical High-Energy Laser or THEL weapon system will provide an innovative solution for the acquisition and close-in engagement problems associated with these threats not offered by existing systems and give the air defense commander an option that he does not currently have. It will significantly enhance the force protection for combat forces and theater-level assets for the Force XXI Army.

For the past several years, SSDC has been pursuing this concept that could provide new air and missile defense capability in force protection missions. Numerous Department of Defense high-energy laser development programs over the last 20 years have proven and demonstrated the laser beam generation and beam-pointing technologies for the THEL concept. Force XXI advancements in the area of real-time situational awareness now make it possible to capitalize on the attributes of a THEL in operational scenarios.

A THEL will be able to rapidly fire for close-in engagements where timelines are very short and cost only a few thousand dollars per kill or less with a deep magazine to counter saturation attacks. Not only can THELs destroy, but they can also degrade, disrupt, or damage to enhance operational flexibility and effectiveness against a wide variety of air threats. This system can, therefore, operate as an effective new weapon node in the short- to medium-range air defense architecture.

The effectiveness of high-energy lasers against short-range rockets was tested and demonstrated in the "Nautilus" program, an outgrowth of Project Strong Safety, in collaboration with Israel. The program was conducted primarily at SSDC's High-Energy Laser Systems Test Facility or HELSTF at White Sands Missile Range, N. M., and utilized a fraction of the power of the HELSTF Mid-Infrared Advanced Chemical Laser (MIRACL) to emulate the THEL weapon concept performance.

The MIRACL is a megawatt-class deuterium fluoride chemical laser that has been operational at HELSTF since the early 1980s. After a series of static and dynamic tests, the program successfully destroyed a short-range rocket in flight on February 9, 1996. This success triggered the next development in the SSDC THEL effort.

In April 1996, the Prime Minister of Israel, Shimon Peres, met with President Clinton and Secretary of Defense William Perry. During the meeting, the United States made a commitment to assist Israel in the development of a THEL Advanced Concept Technology Demonstrator by the end of 1997, based on the success of the Nautilus program, to defeat the threat posed by Katyusha and other short-range rockets against the cities in northern Israel.

In July 1996, a contract was awarded by SSDC to TRW, Inc., of Redondo Beach, Calif., for the design, development, and fabrication of a THEL demonstrator, which will be a transportable, tactical-sized deuterium fluoride chemical laser. Approximately 18 months will be required to design and build the system, followed by 12–18 months of testing.

If successful, the demonstrator may pave the way for future development of a THEL for use in US peacekeeping/contingency operations. The US Army Air Defense Artillery School in Fort Bliss, Texas, already officially designated as the proponent for THEL by the Training and Doctrine Command, is planning to develop a mission need statement and operational concept that could lead to an operational requirement for a THEL system.

Evolving high-energy laser, beam control, and digital battlefield information technologies promise to combine into a highly effective force protection THEL weapon system for the Force XXI Army.

Prepared September 1996 by John J. Wachs, Weapons Directorate, MDSTC, SSDC High-Energy Lasers Weapons

4.10 Lasers for Air Defense

In its 1984 directed energy plan, SDIO planned to develop an acquisition, tracking, pointing, and fire control (ATP/FC) subsystem for directed energy weapons by fiscal year 1990 for $1298 million. Through fiscal year 1993, SDIO allocated $1634 million to this program, accomplishing some but not all of the program objectives. SDIO estimated that it will cost $180 million and take 3 years to resolve the majority of the remaining technical issues. For another $100 million, the ATP technology could be demonstrated in space.

All directed energy weapons need an ATP/FC system. In general terms, the system must quickly engage a large number of targets by placing a directed energy beam on the aimpoint of each target. These time and accuracy constraints dictate a rapid succession of handovers from one sensor to another. Each successive sensor in the system has a smaller field of view and greater accuracy.

The system locks onto the infrared signature of a missile (acquisition), calculates the flight path of the missile (tracking), calculates an aimpoint on the missile and directs the beam to the aimpoint (pointing), and assesses the results and selects the next target (fire control). Depending on the mission of the directed energy system, the ATP/FC system must perform these functions when ballistic missiles are in their boost, post-boost, and/or midcourse phases of flight.

The basic goal of the program was to resolve the technical issues sufficiently to support a space test of a directed energy weapon by 1990. The overall technology performance objectives in the 1984 plan were as follows:

- Reduce the effect on the accuracy of pointing and tracking devices of vibrations caused by operation of the spacecraft and laser to less than 4 in. on the target.
- Develop the capability to rapidly retarget the laser beam from one target to another in less than 2 s.
- Develop the capability to track targets at ranges of 2600–3100 miles at an accuracy of about 4 in.
- Develop fire control computer software to handle more than 100 targets at a rate of more than one target per second. The fire control functions are missile plume to missile hard-body handover, tracking of multiple targets, target identification, aimpoint selection, and damage assessment.

The plan specified that $1298 million would be required from focal years 1986 through 1990 to develop the system components and to fly space experiments to resolve integration and space operation issues. Experiments would permit the space test of a directed energy weapon in 1990.

SDIO met the plan's objectives for pointing and tracking technology and rapid retargeting technology for directed energy weapons. It did not meet the objectives for developing long-range fine tracking and fire control software. While not meeting all objectives, SD10 believes it has met the basic program goal of resolving technical issues sufficiently to support a space test of directed energy technology. Through fiscal year 1993, SDIO spent about $1684 million developing ATP/FC technologies. This amount is about $286 million more than what SD10 estimated was needed to accomplish the objectives. A majority of the funding was spent on a series of space- and ground-based experiments. All major space tracking experiments were canceled before completion due to a lack of funding. However, two space pointing experiments were completed.

At a cost of about $262 million, SDIO reported that it completed the Relay Mirror Experiment and the Low-Power Atmospheric Compensation Experiment, which were focused on resolving issues related to the ground-based laser program. Each was placed in a separate orbit by one Delta booster in 1990. The Relay Mirror Experiment successfully demonstrated high-pointing accuracy, laser beam stability, and long-duration beam relays. The Low-Power Atmospheric Compensation Experiment successfully demonstrated low-power technology to compensate for laser beam distortions, which occur when beams go through the atmosphere from ground to space.

SDIO had spent about $684 million from fiscal years 1985 through 1991 planning, designing, and fabricating hardware for four ATP/FC space experiments that were canceled before completion for the following reasons:

- Talon Gold was intended to demonstrate precision tracking and pointing in space for targeting satellites and boosters. After spending about $26 million on Talon

Gold, SDIO canceled the experiment because the cost estimates for integration and launch had increased an additional $600 million.

- Pathfinder was started in September 1986 and was canceled in 1987 because it was too expensive. SD10 had spent about $40 million on this experiment, which was to address plume phenomenology using a sensor array on the space shuttle.
- The Starlab space experiment was intended to demonstrate precision tracking and would have used the space shuttle to accomplish the experiment. After spending about $603 million developing Starlab, SDIO canceled this experiment in part because the Challenger accident led to nearly a 3-year delay in the launch date, greatly increasing the overall cost. This coupled with changing priorities in the directed energy program led to changes in requirements and increased costs, which made the experiment too expensive to complete.
- Altair, which was canceled after SDIO had spent about $16 million in development costs, was intended to demonstrate the same types of technologies as Starlab, and was planned to use some of the hardware developed for Starlab. An SDIO official estimated that it would have cost $330 million to complete Altair.

SDIO replaced the Altair space experiment with a nonspace ATP/FC experiment called High-Altitude Balloon Experiment. This experiment was intended to achieve most of the same objectives as Altair but at a much lower estimated cost of $76 million. Balloons were used to carry AW/FC devices to an altitude of about 30 km where these devices will be used to acquire and track missiles in the boost phase. SDIO's program manager for ATP/FC expected this experiment to yield from 80 to 90 % of the data that would have been obtained from a space experiment.

SDIO designed and constructed a Rapid Retargeting/Precision Pointing [R2P2] simulator that emulated the dynamics of a large spacecraft (e.g., motion and vibration). Using this facility, SDIO developed and tested techniques for ensuring the stability, accuracy, and precision of a simulated directed energy weapon's pointing device under rapid retargeting situations. This project demonstrated, within the limits of a ground laboratory, that ATP/FC techniques should work in space at the levels established in the original program plan. SDIO will have spent about $42 million on this project from fiscal years 1986 through 1993.

Two other projects also demonstrated ATP/FC techniques. The Space Active Vibration Isolation project developed and tested ATP/FC techniques for negating the effects of spacecraft and weapon vibrations on the pointing device. This project produced hardware and technology that have improved the pointing stability of directed energy devices to below the program goal of less than 100 nanoradians or about 4 in. from a distance of 1000 km. This project was followed by the Space Integrated Controls Experiment, which improved the pointing stability even further. SDIO spent about $37 million on these two projects from fiscal years 1986 through 1993.

As of 1993 SDIO estimated that it would cost $180 million and take three more years to resolve the vast majority of the ATP/FC technical issues and perform integrated ATP experiments against real targets from the High-Altitude Balloon

Experiment platform. This would substantially complete the objectives of the 1984 plan. An additional $100 million would be needed to demonstrate operation in space, assuming that it would be done as part of another directed energy space experiment such as Star LITE, the experiment planned for the chemical laser. The major technical issues to be resolved from 1993 through 1996 included long-range fine tracking, fire control, integrated ATP/FC, and additional concept development.

For long-range fine tracking, the Solid-State Baser Radar Source program produced two laser illuminators. They still need to be tested in realistic target environments to determine their effectiveness in changing conditions and against a wide variety of targets. In addition, their capabilities must also be developed to support aimpoint selection and maintenance and damage assessment.

Fire control decision software had been demonstrated in computer simulations, but its practicality and robustness had yet to be tested in an integrated field operation. Each of the individual fire control decision algorithms needs to be tested with several sets of scene conditions with real data. Functional integration with sensors and autonomous operation must also be demonstrated. SDIO plans to test the operation of the software on the High-Altitude Balloon Experiment platform against boosting targets at the White Sands Missile Range Test.

4.10.1 Target Acquisition for Combat Operations

This section will be discussing target acquisition, tracking, and combat operation by directed energy weapon systems.

4.10.2 Overview

The goals of this sub-thrust can best be described in three major categories:

1. The measurement of atmospheric parameters on space and time scales required to support Air Force and directed energy weapon system and its missions in Global Reach–Global Power
2. The prediction of the future evolution of the atmosphere from a few hours to a few days
3. The assessment of weather impact on Air Force and directed energy weapon systems and operations

4.10.3 Description

Theater Specification for Dominant Maneuvers sub-thrust conducts R&D programs to better understand the physical and dynamic processes of the lower atmosphere in order to design, develop, test, and transition remote sensing instrumentation, retrieval algorithms, and models in support of air and space warfighters.

Real-Time Measurements of Atmospheric Parameters from Satellites: The measurement goal emphasizes the application and interpretation of satellite sensor data, but also includes in-situ sensor development as part of the program. Measurement emphasis is on the three-dimensional determination of cloud cover at the highest spatial and temporal resolution to develop the models to support global surveillance and tactical warfare.

Atmospheric Optical Turbulence Measurements and Modeling: The technical objective is to specify and predict the atmospheric optical turbulence degradation to ground-, air-, and space-based laser systems. Optical turbulence is highly variable from site to site, season to season, and day to day; however, as yet there are no good predictive models. In a program initiated in FY96, supported by the airborne laser (ABL) SPO, Phillips Lab is obtaining optical turbulence data in theaters of interest to the SPO.

Tactical Remote Sensing: The technical objective is to provide from aircraft and satellite platforms remote sensing of battlefield gases and emissions. The PL/GPOR LIDAR remote sensing program is set up to measure several atmospheric boundary layer parameters in high spatial resolution, including wind profiles, aerosol content, and size distribution, as well as the detection of trace elements, both natural and man-made. Current assets include the following:

- 10.6-μm CO_2 Doppler range-resolved wind profiler, dual tunable 9–11-μm CO_2 Doppler range-resolved wind profiler, and DIAL system to obtain range-resolved water vapor and trace gas profiles
- 1.574-μm portable, environmental, eye-safe LIDAR (or LADAR) for range-resolved aerosol and cloud profiles in addition to cloud depolarization features

New assets soon to be on hand include a coherent, tunable UV DIAL system for water vapor and ozone detection, a mechanical turbulence LIDAR or LADAR for wind shear detection, and a refractive turbulence remote sensor to measure turbulence along a path.

The Theater Forecast for Precision Engagements FTA develops tailored weather products to support combat mission planning and execution worldwide.

Prediction of Clouds and Severe Weather: The prediction goal emphasizes the data fusion of satellite and indigenous data sources in the battlefield where data denial may be a factor. The resulting analysis fields will then be tested in theater-scale prediction models on time and space scales appropriate to tactical weapon delivery.

Virtual Weather: The goal of this program is to produce computer simulations of the atmosphere, with high physical fidelity, valid as a function of location, season, time of day, and geometry. In order to address these requirements, simulation models must be developed that produce three-dimensional structure—the major shortfall of current capability. Initial efforts involve improving the physical reality of current cloud simulation models. Other projects include development of rain, fog, wind, humidity, lightning, and turbulence models. A strong emphasis is placed on physically correct visualization and the generation of radiometrically correct atmospheric scenes.

Weather Impact Decision Aids (WIDAs): The WIDA program is developing software technology for operationally predicting the impact of weather on the performance of airborne electro-optical navigation and weapon targeting systems. The program presently has four major ongoing and/or planned components: (1) Night Vision Goggle Operations Weather Software (NOWS), which will predict the impact of weather on night vision goggle detection range for AFSOC and ACC; (2) IR Target Scene Software (IRTSS), which is developing software that will determine the impact of weather on air-to-ground target scenes in the infrared and produce an IR scene visualization for transition to the Air Force Mission Support System (AFMSS); (3) Weather Automated Mission Planning Software (WAMPS), a new start in FY97 to develop software to automatically incorporate weather impacts on airborne EO systems during theater mission planning in Theater Battle Management Core Systems (TBMCS); and (4) Target Acquisition Weather Software (TAWS), a new start in FY97, will provide a major upgrade to the current operational Electro-Optical Tactical Decision Aid (EOTDA) used by Air Force Weather (AFW) support personnel to provide weapon lock-on and acquisition ranges for EO weapon systems used by ACC:

1. **User Impact**:
   None.
2. **User Impact**:
   Concept/technology.
3. **Images**:
   None.
4. **Related Initiatives**:
   - **Target Acquisition for Combat Operations** (see Sect. 5.1).
   - **Target-Background Discrimination for Surveillance** (see Sect. 5.2)
5. **Related Requirements**:
   None.
6. **Related Categories**:
   Optical Surveillance Effects and Battle-Space Operations
   The importance of maintaining a reduced nuclear force and the emerging conventional ballistic force as a combat and cost-effective weapon is recognized by Air Force Space Command (AFSPC) mission area plans and Space and Missile Systems Command (SMC) development plan having the technology needs in advanced guidance technologies and astrodynamics.
   - **The goals of advanced guidance are:**
     - Global Positioning System (GPS) Range Standardization/Safety Technology
     - Development of new miniature systems to lower range costs 30 % by replacing radar systems and enhance safety with greater accuracy and reliability by FY01
     - GPS accuracy enhancements

- To increase cost effectiveness, missile navigation, and testing accuracy with improved GPS/INS coupling by FY98
- Development of precision fiber optic gyroscope (PFOG) with low-loss integrated optics and fiber couplers and flexured mass accelerometer (FMA) with open-loop, two back-to-back microwave resonant cavities by FY99
- To decrease the reliance on high-cost, high-precision inertial measurement systems with micromechanical updates on accelerometers and gyros by FY01
- To fly the Missile Technology Demonstration III (MTD III) during FY01 to gain data on multiple penetrator warheads delivered on an ICBM
- To develop anti-jam antennas
- To integrate plasma physics with design
- To develop and test materials for antenna windows

- **The goals of astrodynamics are to:**
  - Improve differential correction (DC) accuracy 90 %.
  - Improve propagation accuracy at the end of the prediction period by 90 %.
  - Demonstrate integrated performance of high-accuracy lasers and astrodynamics algorithms to precisely locate and illuminate spacecraft.
  - Demonstrate next-generation initial orbit determination, DC, and propagation for space surveillance.
  - Show deficiencies in current operational DC and propagation which could be eliminated.
  - Demonstrate capability to maintain independent high-accuracy catalog of selected satellites (20–30 objects).

4.11 Target-Background Discrimination for Surveillance

The following sections will apply for this purpose.

4.11.1 Overview

Space-based surveillance, tracking, and interceptor systems must accurately and reliably discriminate target IR and optical signatures from the atmospheric and celestial IR and optical emissions against which the targets are viewed. When sensor specifications are optimized against realistic, accurate simulations of the atmospheric and celestial emissions in the sensor's field of view, then operational performance is significantly enhanced, and system overdesign is minimized, thereby making DOD space-based systems much more affordable. The sub-thrust goals are to develop and demonstrate integrated background clutter mitigation

technologies for SBIRS-High and SBIRS-Low and next-generation hyperspectral surveillance and threat warning systems and to integrate clutter suppression technologies into hardware simulators to provide high-resolution spectral and spatial scene data of atmospheric, cloud, terrain, and celestial background clutter to support system designs.

This sub-thrust:

1. Defines the impact of optical and infrared backgrounds on surveillance and threat warning, theater and national missile defense, and intelligence, surveillance, and reconnaissance systems
2. Measures atmospheric and celestial infrared, ultraviolet, and visible backgrounds using satellite and rocket-borne sensors
3. Models atmospheric and celestial optical and infrared backgrounds for the full range of operational conditions and system design trade space
4. Provides real-world background scenes, global background statistics, and reliable background scene models to support system engineering trade studies
5. Measures and models in-flight infrared signatures of aircraft and missiles
6. Distributes and provides online access to background phenomenology data
7. Defines optical and infrared background requirements for surveillance systems and battle-space simulations

4.11.2 Description

This sub-thrust defines the impact of optical and infrared backgrounds on surveillance and threat warning, theater and national missile defense, and intelligence, surveillance, and reconnaissance systems and measures atmospheric and celestial infrared, ultraviolet, and visible backgrounds using satellite and rocket-borne sensors. The two main project areas of this sub-thrust are backgrounds and target phenomenology and background clutter mitigation.

The goals of the backgrounds and target future technical architecture (FTA) are to provide high-resolution spectral and spatial scene data of atmospheric, cloud, terrain, and celestial background clutter to support SBIRS designs; provide high-throughput data processing, analysis, and distribution of background phenomenology data for SBIRS, BMDO, and other DOD programs; measure in-flight infrared signatures of aircraft and missiles; and develop models capable of predicting the infrared characteristics of all aircraft, particularly in the design and development stages. A major task under this FTA is to construct infrared atmospheric and celestial background scenes from the superb data measured by the MSX satellite and atmospheric scenes from the highly successful MSTI-3 satellite mission. The combined data set will be used to upgrade models for the design of the new SBIRS surveillance system.

The goals of the background clutter mitigation FTA are to (1) characterize and predict atmospheric, cloud, and terrain infrared background clutter for the full range of SBIRS operational conditions and system design trade space and (2) assess the

impact of background clutter on SBIRS performance and mission capabilities. Spatial and temporal structure in atmospheric, cloud, and terrain backgrounds produces clutter against which infrared and optical sensor systems must detect and track theater ballistic missiles, cruise missiles, and aircraft threats as well as perform technical intelligence missions.

Under this program, tasks are being performed to:

1. Provide background clutter codes to extrapolate measured background data to the full SBIRS design trade space and all SBIRS operational conditions.
2. Provide background model uncertainty bounds.
3. Assess and upgrade background clutter codes for SBIRS using MSX and MSTI-3 data.
4. Develop dynamic, statistical background clutter models to support adaptive hyperspectral imaging.
5. Provide expert user interface for code applications:

1. **User Impact**:
   None.
2. **User Impact**:
   Concept/technology.
3. **Images**:
   None.
4. **Related Initiatives**:

   - **Space-Based Infrared System—Low Earth Orbit (SBIRS-Low)**
     The space-based infrared system (SBIRS) is in response to the US military forces increasing the need for accurate and timely warning of tactical missile attack. SBIRS will replace the current Defense Support Program (DSP) designed to meet US infrared space-based surveillance and warning needs through the next two to three decades. SBIRS improves support to theater CINCs, US deployed forces, and allies by providing detailed information in the four mission areas of missile warning, missile defense, technical intelligence, and battle-space characterization. SBIRS will provide significant performance enhancements over DSP by improving quality and timeliness of missile warning data. SBIRS should enhance information superiority and support the Joint Vision 2010 operational concepts of full-dimensional protection and precision engagement, by providing this data directly to theater commanders in a timely, survivable manner, thus enabling US forces' immediate reaction to threat.

     The SBIRS space segment includes a high and low component. The high component comprises six satellites: four in geosynchronous (GEO) Earth orbit and two hosted payloads in highly elliptical orbit (HEO). The low component includes approximately 24 low Earth orbit (LEO) satellites. The SBIRS high component will meet a subset of the operational requirements, including all key threshold requirements. The SBIRS low component will provide a unique, precision, midcourse tracking capability critical for

effective ballistic missile defense, as well as enhanced capability in support of other SBIRS missions. SBIRS-High, complemented by SBIRS-Low satellites, will meet all of its operational requirements.

The SBIRS ground segment includes a continental US CONUS-based Mission Control Station (MCS), a MCS backup (MCSB), a survivable MCS (SMCS), overseas relay ground stations, Multi-Mission Mobile Processors (M3Ps), and associated communication links. The SBIRS ground segment will be delivered incrementally. The first increment, scheduled to be operational in FY99, consolidates DSP and Attack and Launch Early Reporting to theater ground stations into a single CONUS ground station and will operate with DSP satellite data. The second increment, scheduled for FY02, will provide the necessary ground segment functions required for the new high-altitude SBIRS satellites and the residual DSP satellites. Included in the second increment will be mobile terminals capable of fulfilling the Army Joint Tactical Ground Station in-theater and SBIRS strategic processing requirements. A third increment, which will be operational in FY03, will add the necessary ground segment functions for the first LEO satellite scheduled to be deployed in FY04.

Background Information:

SBIRS was initiated in 1995 as a replacement for the Follow-on Early Warning System acquisition, which was canceled due to cost and requirement problems. Since SBIRS satellites need to be completed before the last DSP satellite is launched, it was placed on an accelerated schedule and selected as a lead program for acquisition reform. Much of the traditional required documentation was reduced or consolidated into a Single Acquisition Management Plan, and emphasis was placed on direct involvement through Integrated Product Teams (IPTs) rather than traditional documentation reviews.

The SBIRS high component entered the EMD phase following a Milestone II DAB review in October 1996. This decision was supported by an OA conducted by AFOTEC and reviewed by DOT&E.

The first phase of IOT&E will be conducted in 1999 to verify the performance of the Increment 1 ground station. Due to the critical role SBIRS plays in Integrated Tactical Warning and Attack Assessment (ITW/AA) of attack on the CONUS, DOT&E has become involved in this program early. DOT&E works closely with AFOTEC, the program office, and all users to ensure that the acquisition strategy fosters an operationally effective and suitable system while maintaining cost effectiveness. DOT&E has supported SBIRS acquisition reform through heavy involvement in IPTs, early involvement in combined developmental and operational tests, and consolidation of developmental and operational test plans into a single integrated T&E plan.

The SBIRS test program includes a combination of OAs, combined DT/OT testing, and dedicated IOT&E. These OT&E events will progress in a building block manner beginning with analyses, modeling, and validated

simulation and ending with hardware-in-the-loop (HWIL) test-beds and field tests. Modeling, simulation, and test-beds will be used to assess those areas in which field testing cannot be conducted, such as actual missile attacks and operation in nuclear environments. SBIRS operational effectiveness and suitability will be assessed on the basis of IOT&Es of each of the three major increments, which will include fixed and mobile assets.

Test and Evaluation Activity:

In 1998, DOT&E approved an initial TEMP that defined the top-level test strategy and mapped it into the overall acquisition strategy. DOT&E also continued its oversight of the following areas (each of which could impact schedule, cost, and system performance):

- Progress toward Increment 1 IOT&E
- HWIL test-bed definition and dynamic effect modeling for Increment 2
- Risk reduction efforts for Increment 3
- Testability of SBIRS/NMD requirements for Increment 3

Progress toward Increment 1 IOT&E was assessed by an AFOTEC (OA) that addressed four areas: (1) major issues potentially affecting effectiveness and suitability, (2) programmatic voids, (3) testability of user requirements, and (4) ability of the program to support operational testing. DOT&E has specific program concerns about SBIRS Increment 1: (1) immature ground system software and delays in requirement performance verification, (2) delays in procuring high-reliability communication links from the overseas ground stations to the Mission Control Station, and (3) adequate hardware for crew training. The program office is addressing many of these findings through specific risk reduction efforts to ensure readiness to enter IOT&E for the Increment 1 ground system in April 1999. There are still testability concerns involving the difficulty of testing SBIRS/NMD operational requirements within an acceptable confidence limit.

Test and Evaluation Assessment:

Year 2000 (Y2K) testing for SBIRS is well underway, and there are no anticipated problems with the system. Due to extensive use of commercial software and close cooperation between the contractor team and the Air Force, an adequate verification program is in place. Final Y2K testing will be complete prior to the start of IOT&E for the Increment 1 ground station in April 1999.

The major near-term challenge for the SBIRS program is to ensure a seamless transfer of operations from the current DSP ground stations to the new SBIRS Increment 1 MCS. This demanding task is complicated by the compressed timeline and issues associated with shared use facilities at the overseas relay ground stations. Additionally, there have been significant delays in validating software performance of the Increment 1 ground system. Other near-term challenges for the SBIRS program include the adequacy of test-bed design and the scope of models and simulations needed to validate the stressing requirements for the SBIRS-High satellites and MCS Increment

2 and the significant technical risks associated with accelerated deployment of the low component by FY04. The demanding SBIRS-High requirements are a significant improvement over DSP's demonstrated performance and require extensive testing to validate and assure the system's performance. For HWIL test-beds, continued attention must be given to ensure that the test-beds are adequate to support OT&E, including the need to portray dynamic backgrounds that accurately portray the Earth's background as seen from space.

DOT&E's assessment is that the SBIRS compressed schedule to achieve "online in '99" remains high risk, and delays in software integration and testing pose an increased risk in a "zero margin" schedule leading to Increment 1 IOT&E scheduled for April 1999. The primary challenge for Increment 1 is the verification of software performance and reliability. There have been significant delays in verifying software performance and reliability, as well as delays in hardware installation at the remote ground stations. While this type of problem is not unusual, many systems interfacing with the SBIRS MCS are 1970's legacy reporting systems, whose interfaces may not be adequately documented. Delays in starting the testing of these interfaces put an inordinate amount of pressure on the first opportunity success. The "never fail" nature of ITW/AA systems requires extensive "online" testing to validate reliable Increment 1 operations and a period of parallel operations prior to declaration of IOC. Any significant delays to IOT&E would lead to "ripple effect" delays in the Increment 1 IOC date and further delay the IOC dates for subsequent ground system increments. Also, there is a concern that the SBIRS Increment 1 ground system includes voids in areas of fault detection and isolation, operator training, and manpower.

SBIRS Increment 2 (both space and ground elements) remains on schedule, but faces continued challenges in the areas of simulation and test-bed development. For Increment 2, progress has been made in identifying real-world, dynamic effects in the short and medium wavelengths detected by the greatly improved SBIRS-High sensors. The operational impact of these effects must be quantified and the SBIRS-High sensor design shown to be robust enough to handle these natural phenomena. The resolution of these issues can be best accomplished by incorporation of adequate testing processes into the baseline sensor ground-testing program. Until this testing is completed, the capabilities of the SBIRS sensor and signal processing to operate in the space environment remain a major concern.

Continuing significant technical problems with the SBIRS-Low, PDRR satellites demonstrate the wisdom of an extensive PDRR test phase before entering EMD to start the construction of operational SBIRS-Low satellites. The current schedule of events is very compressed and does not allow full evaluation of the PDRR satellites' performance. The current baseline SBIRS-Low schedule requires successful completion of many difficult activities proceeding in parallel toward a successful FY04 first launch, thus violating recommendations outlined in the recently completed Welch report on missile

defense systems. Any additional delays in the PDRR competing contractors' programs will require starting the EMD prior to completion of PDRR to meet the congressional goal of an FY04 first launch. DOT&E is concerned that the baseline schedule, which includes the Flight Demonstrations System and the Low Altitude Demonstrations System, will be delayed, presenting very few opportunities to collect "real-world" performance data on contractor designs to assess their ability to meet draft performance requirements. This period of evaluation of PDRR results is critical since "lessons learned" from PDRR test activities form the foundation for the government and contractor teams to perform cost as an independent value satellite design trades. Any significant problems encountered during the PDRR phase (given the compressed schedule) may lead to premature launching of inadequately designed and tested satellites to maintain the FY04 initial deployment date.

To support Milestone II decisions, DOT&E has worked closely with the AFOTEC, the program office, and the user community, to ensure that the acquisition strategy throughout the acquisition cycle fosters an operationally effective and suitable system, while maintaining cost effectiveness. This early involvement included active membership in IPTs, fostering combined developmental and operational tests, early validation of software maturity, and consolidation of developmental and operational test plans into a single integrated T&E plan.

5. **Related Requirements**:
   None.
6. **Related Categories**:

- **Contributing Sensors**
  These sensors provide observation data on satellites to USSPACECOM on a contributing basis, but are not directly under the operational control of USSPACECOM. Both mechanical radars and electro-optical systems are included in this category.
- **Satellite Operations**
  The DoD procures, operates, and maintains a myriad of satellite systems to support national and tactical communications, missile warning, nuclear detonation detection, navigation, weather, and environmental monitoring. DoD's satellite operations include the Defense Satellite Communications System (DSCS), Milstar, Fleet SATCOM (FLTSATCOM) system, UHF Follow-on (UFO) system, the Defense Support Program (DSP), the Nuclear Detonation Detection System (NUDET), the Global Positioning System (GPS, also referred to as POSNAVTIME or position, navigation, and time), and the Defense Meteorological Satellite Program (DMSP) systems that are included in this category.
- **Space-Based Warning System**
  This category addresses space systems and sensors that have a surveillance and warning mission, which are operational, in development, or being studied.

- **DoD Space Surveillance Programs**
  A constant and vigilant surveillance of potentially hostile military threats is critical in preserving the operational effectiveness of our armed forces around the world. Naval Space Command manages two distinct surveillance efforts in support of Fleet and Fleet Marine Forces: tracking satellites in orbit and monitoring over-the-horizon threats from sea and air forces.

  Over one million satellite detections, or observations, are collected by this surveillance network each month. Data gathered is transmitted to a computer center at Naval Space Command headquarters in Dahlgren, where it is used to constantly update a database of spacecraft orbital elements. This information is reported to Fleet and Fleet Marine Forces to alert them when particular satellites of interest are overhead. The command also maintains a catalog of all Earth-orbiting satellites and supports USSPACECOM as part of the nation's worldwide Space Surveillance Network.

References

1. Weichel H (1990) Laser beam propagation in the atmosphere, SPIE. Optical Engineering Press, Bellingham
2. Kopeika NS (1990) A system engineering approach to imaging, SPIE. Optical Engineering Press, Bellingham
3. Andrews LC, Phillips RL (2005) Laser beam propagation through random media, 2nd edn. SPIE Press, Bellingham

Chapter 5
Laser-Directed Energy Concepts

This chapter will discuss directed energy concepts for strategic defense. We will talk about defensive weapons as a countermeasure against any measure that is applied in terms of a lethal weapon against friendly targets. Directed energy concepts can play unique roles in strategic defense because of their reaction time, speed of light engagement, and large geographic converge. This chapter discusses the main directed energy concepts, engagements in which they could have significant advantage, and their expected performance in them. It covers both boost-phase engagements and midcourse applications and contrasts these results with those of earlier analyses (Fig. 5.1).

5.1 Laser Beam and Material Interactions and Its Lethality

Weapons are devices, which deliver sufficient energy to targets to damage them. Weapon design involves a dialog between weapon designers and military planners. Designers create means of projecting energy, and planners have targets that they would like to destroy. Effective design requires knowledge of the targets and the circumstances of their engagement, and effective planning requires knowledge of the weapons and their characteristics. However, in new and emerging areas of weaponry, designers and planners often do not speak the same language. As a result, designers can operate in ignorance of operational realities, and planners can assume that anything involving new technology will meet all their needs.

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_5

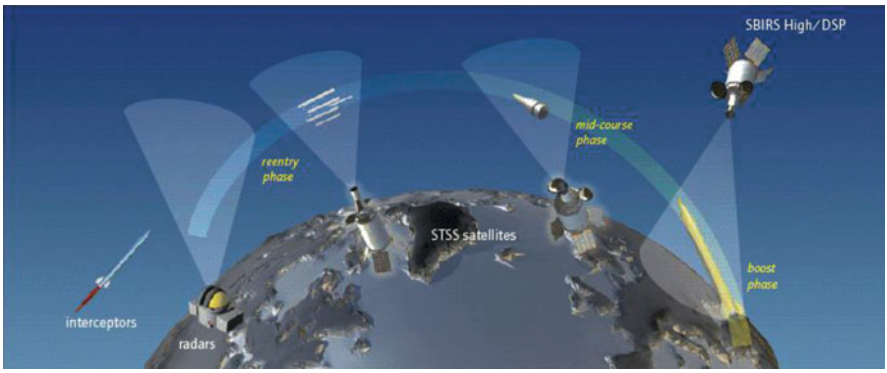
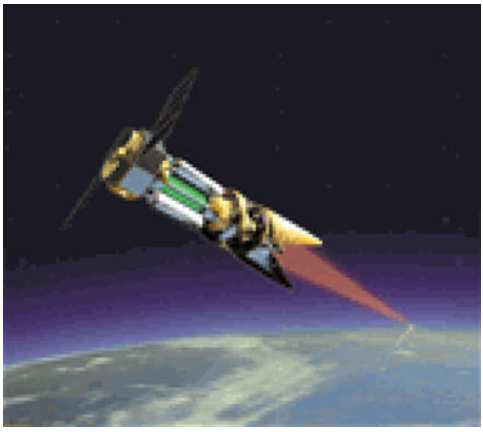


Fig. 5.1 National configurations of space tracking surveillance system (STSS) and space-based infrared system (SBIRS)



Directed energy weapons are no different. However, while there are books and manuals that deal with the issues affecting the utility of nuclear missiles and rifles, there is no comparable source of information for directed energy weapons. I have tried to fill that void with this book as well as dealing with technical and theory of interaction of high-energy laser in particular and avoid aspect of high-power microwave in general as lethal weapons.

This book is way to explain high-energy laser power and how directed energy weapons work, how the energy of high-power laser as weapon is propagated to the target, and how the weapons/laser beam–target interaction creates effects (damage) in the target. The mathematics and physics of effects of high-power laser radiation beam interact with its targets and analyzes the damages within a period of dual time.

This is a technical exposition, written at the undergraduate physics and engineering level that could serve either as a textbook or as a reference text for technical practitioners. The text touches upon kinetic energy weapons in addition to high-power lasers and stays away from the other two aspect beam weapons, namely, microwaves and particle beams. Numerous unclassified articles and literature both

on high-power microwaves and particle beams can be found on the Internet that are published by experts in these **two fields xx and yy**. A directed energy weapon is a type of weapon that emits energy in a particular direction by a means other than a projectile. It transfers energy to a target for a desired effect. Some of these weapons are real or practicable; some are science fiction. The energy is in various forms:

- Electromagnetic radiation (typically lasers or masers).
- Particles with mass (particle beam weapons).
- Fictional weapons often use some sort of radiation or energetic particle that does not exist in the real world or where the physical nature of the energy and its means of transmission are not detailed and the visible effects would be impossible in the real world.

Some of these weapons are known as death rays or ray guns and are usually portrayed as projecting energy at a person or object to kill or destroy.

Some lethal directed energy weapons are under active research and development, but most examples appear in science fiction (or nonfunctional toys and film props).

5.2 Introduction to Effectiveness of Directed Energy Weapons

The effectiveness of a defensive weapon system is measured by its ability to deny the attacking system succession accomplishing its mission.

Lethality of a directed energy weapon (DEW) is, in the simplest terms, its capability to destroy a target and incoming threat. It is appropriate to speak of lethality as the capability of directed energy weapons to prevent a target from accomplishing a particular mission. This requirement could be fulfilled in the form of “hard kill” (physical destruction of the warhead) or “soft kill” (mission impairment).

Further study of the “kill” phenomenon requires a distinction between immediate (within milliseconds) and delayed kill. The difference results in tracking the assets wasted and the unnecessary dilution of defensive systems in a time-critical situation [1].

To meet the kill requirements, weapon systems using three different energy sources have been proposed. Historically, chemical and nuclear explosives have been used. Together, they form the "potential energy" weapon (PEW) groups. More recently, kinetic energy weapons (KEWs) and directed energy weapons (DEWs) have been proposed, and engineering development has begun.

Thus, there may be several measures of lethality for a given target set. For an incoming missile threat, one may define lethality criteria to structure damage of the target in terms of hard kill and other criteria relating to destruction or indefinite interruption of the sensors or guiding system on which the missile depends to

accomplish its function. Similarly, actual destruction of a booster or reentry vehicle sets certain lethality criteria, but methods of destroying accurate weapon delivery, such as destruction of guidance electronics, in terms of soft kill, may also generate acceptable lethality criteria for system designers. However, in the latter case, verification of a kill becomes problematical.

The ability of laser beams (pulsed, continuous wave (CW), and repetitively pulsed), at infrared, visible, ultraviolet, and X-ray wavelength, to destroy various targets is analyzed. First, the physics of the interaction of various laser beams as part of directed energy weapon with materials is examined. This information is used to assess the effect of a given incident power or energy flounce on the target and the ability of the target to perform its mission after such an attack. The arguments, then, are used to size weapon system to destroy enemy targets (lethality) and related safety factor and issues in regard to safety of friendly troops present in the battlefield of such engagements.

The fundamental kill mechanism of CW or quasi-CW repetitively pulsed laser beam is heating, with subsequent melting and/or evaporation of the wall of a liquid or solid booster rocket. Subsequently, ignition of booster fuel may take place, or mechanical failure of structures may occur before completion of burn through. In a similar manner, the wall of the bus and components inside it may be damaged, so that the intended function of the missile is thwarted.

In addition to energy deposition, momentum is also transferred to the target by directed energy beams. Momentum transfer can damage targets through mechanical shearing or buckling. This damage mechanism has been demonstrated by pulsed laser beams for pulses less than or equal to 2 μs. Kill through repeated impulse damage has system-level advantages over thermal kill since the pointing requirements are far less severe. Momentum transfer may also be used as a discrimination tool for reentry vehicles and decoys in the midcourse such as technique employed in LIDAR (laser radar) applications. The interaction of CW or quasi-CW laser radiation with targets is discussed, and lethality criteria for this type of beams are derived. Finally, this chapter summarizes the main conclusions, setting lower limits on power and energy that are imposed on a DEW of laser-type system by the lethality requirements, and as result the measurements that are needed to take under consideration to address the issue of LRST to friendly troops in vicinity of target engagement field.

5.3 The Mathematics of Diffusion

The history of heat conduction and its mathematical theory principally starts with Fourier and was set forth by him in his *Theorie analytique de la Chaleur* with solutions of problems naturally arising from it. While Fourier treated a large number of cases, including most of those we shall see occasional consideration, his work was extended and applied to more complicated problem by his contemporaries Laplace and Poisson and later by a number of others.

One of the most important applications that we can consider here is the mathematical diffusion theory of heat conduction and principle of high-power laser beams and the matter that both have industrial and military applications. This includes applications such as heating, melting, vaporization, and plasma production. These phenomena, which likewise can involve an interaction of a high-power laser beam, are indeed sometimes called laser effects. Some of these considerations will be out of scope of this book, but its principle is presented here, and further investigation in any particular application should be studied by the reader. To introduce some general idea about how to go about these applications and laying out *The Mathematics of Diffusio*n, we start with the diffusion equation and the diffusion process.

5.3.1 The Diffusion Process and Basic Hypothesis of Mathematical Theory

Diffusion is the process by which the matter is transported from one part of a system to another as a result of random molecular motions. The transfer of heat by conduction is also due to random molecular motions, and there is an obvious analogy between the two processes [2]. This was recognized by Fick in 1855, which first put diffusion on a quantitative basis by adopting the mathematical equation of heat conduction derived some years earlier by Fourier in 1822. The mathematical theory of diffusion in isotropic substances is therefore based on the hypothesis that the rate of transfer of diffusing substances through unit area of a section is proportional to the concentration gradient measured normal to the section, i.e.,

$$F = -D\frac{\partial C}{\partial X} \tag{5.1}$$

where F is the rate of transfer per unit area of section, C is the concentration of diffusing substances, x is the space coordinate measured normal to the section, and D is called the diffusion coefficient. In some cases diffusion in dilute solutions D can reasonably be taken as constant, while in others, e.g., diffusion in high polymers, it depends very markedly on concentration [2]. In the case of heat transfer and solving the problem involving the temperature field determination, one should obtain differential heat conduction. In this case the temperature increase along the normal to the isothermal surfaces is characterized by a temperature gradient (grad $\boldsymbol{T}$). A temperature gradient is a vector along the normal to the isothermal surface in the direction of the increasing temperature, i.e.,

$$\text{grad}\ \boldsymbol{T} = \boldsymbol{n_0}(\partial T/\partial n) \tag{5.2}$$

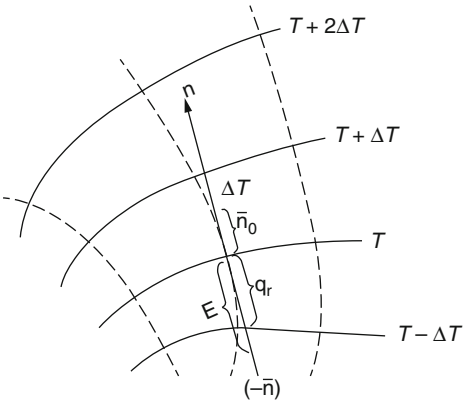


Fig. 5.2 Isotherms of the temperature filed [3] (letters with arrows corresponding to boldface type in the text)

where $\mathbf{n}_0$ denotes a unit vector, along the normal in the direction of the temperature change according to Fig. 5.2 and $\partial T/\partial n$ is the temperature derivative along the normal (n) to the isothermal surface [4].

The gradient is also denoted by ∇.

Gradient components along the Cartesian coordinates are identical to the appropriate partial derivatives, so that

$$\text{grad } \boldsymbol{T} = \vec{\nabla} \cdot \boldsymbol{T} = \boldsymbol{i}\frac{\partial T}{\partial x} + \boldsymbol{j}\frac{\partial T}{\partial y} + \boldsymbol{k}\frac{\partial T}{\partial z} \tag{5.3}$$

where $\boldsymbol{i}, \boldsymbol{j}$, and $\boldsymbol{k}$ are mutual orthogonal vectors of a unit length along the coordinate axes. This relation is possible because any vector may be represented as a vector sum of three components along the coordinate axes.

The concept of a temperature-filed intensity may be introduced as follows:

$$\vec{\boldsymbol{E}} = -\text{ grad } \boldsymbol{T} \tag{5.4}$$

The vector $\vec{\boldsymbol{E}}$ is referred to as a vector of the temperature-filed intensity.

The quantity of heat transferred per unit time per unit area of the isothermal surface is referred to as a heat flux; *the appropriate vector is obtained by the relation*

$$\boldsymbol{q} = (-\boldsymbol{n_0})\frac{\mathrm{d}Q}{\mathrm{d}t}\frac{1}{S} \tag{5.5}$$

where $\frac{\mathrm{d}Q}{\mathrm{d}t}$ the quantity of heat is transferred per unit time or the heat-flow rate, $\mathcal{S}$ is the isothermal surface area, and $(-\boldsymbol{n_0})$ is a unit vector along the normal to the surface area $\mathcal{S}$ in the direction of the decreasing temperature. In this case the vector $\boldsymbol{q}$ is therefore designated as heat flux vector, the direction of which is opposite to that of the temperature gradient (both vectors follow the normal to the isothermal surface,

but their directions are opposite to each other). The projection of the vector **q** on any arbitrary direction is also the vector q_1, the scalar quantity of which is $q\cos(n, l)$.

The lines which coincide with the direction of vector $\boldsymbol{q}$ are referred to as heat-flow lines. These are perpendicular to the isothermal surfaces at the intersection points. A tangent to the heat-flow lines taken in the opposite direction yields the temperature gradient direction (see Fig. 5.2).

The fundamental heat conduction law may be formulated as follows: *the heat flux is proportional to the temperature field intensity*, *or the heat flux is proportional to the temperature gradient*, i.e.,

$$\boldsymbol{q} = k\vec{\boldsymbol{E}} = -\operatorname{grad}\boldsymbol{T} = -k\boldsymbol{n}_0(\partial T/\partial n) \tag{5.6}$$

where k the proportionality factor is also called thermal conductivity. To reveal the physical significance of the thermal conductivity, we shall write the basic relation (Eq. 5.6) for a steady one-dimensional temperature field for the situation where the temperature depends only on one coordinate which is normal to the isothermal surfaces. The scalar quantity of the heat flux vector is

$$q = -k\frac{\mathrm{d}T}{\mathrm{d}x}\left(\frac{\partial T}{\partial x} = \frac{\partial T}{\partial y} = \frac{\partial T}{\partial z} = 0\right) \tag{5.7}$$

If the temperature gradient is a constant value $\left(\frac{\partial T}{\partial x} = \text{constant}\right)$, which means the temperature variation with x follows the linear law, then it may be written as

$$\frac{\partial T}{\partial x} = \frac{T_2 - T_1}{x_2 - x_1} = \text{constant} \tag{5.8}$$

Hence the heat-flow rate $\frac{\mathrm{d}Q}{\mathrm{d}t}$ is also a constant value as

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{Q}{t} = \text{constant} \tag{5.9}$$

where Q is the quantity of heat following in the time τ.

It follows from Eqs. 5.5 and 5.9 that

$$\frac{Q}{St} = -k\frac{T_2 - T_1}{x_2 - x_1} = k\frac{T_2 - T_1}{x_1 - x_2} \tag{5.10}$$

since $T_1 > T_2$ and $x_2 > x_1$.

Thus the thermal conductivity is equal to the heat flowing per unit time and per unit surface when the temperature difference per unit length of the normal is 1°. Thermal conductivity has dimensions of kcal/m h °C or W/m °C. Thermal conductivity is a physical property of a body characterizing its ability to transfer heat. The physical significance of the thermal conductivity and its dependence on basic

properties of a body may be better understood when we consider the heat transfer mechanism in a body in a specific state [4].

The relation $k/(x_2 - x_1) = k/\Delta x$ (kcal/m^2 h °C or W/m^2 °C) is called thermal conductance of a certain portion of a body, and the inverse value $\Delta x/k$ (m^2 h °C) or (m^2 °C/w) is the thermal resistance of this portion of a body, and the magnitude of it varies for different materials over a wide range.

5.3.2 *The Differential Equation of Diffusion Equation*

The necessary condition for heat conduction is the existence of a temperature gradient. Experience shows that heat is transferred by conduction in the direction normal to the isothermal surface from a higher-temperature level to a lower one.

To solve problems involving the temperature field determination, one should obtain a differential heat conduction equation. A differential of heat equation is a mathematical relationship between physical quantities characterizing the phenomenon considered, these quantities being functions of space and time. Such an equation characterizes the physical process at any point of a body at any moment. A differential heat equation provides a relation between temperature, time, and coordinates of an elementary volume. A differential equation will be derived by a simplified method. A one-dimensional temperature field is assumed (heat propagates in only one direction, say, in the direction of the x-axis). Under these conditions we assume that the thermal coefficients are to be independent of spatial coordinates and time. We single out an elementary rectangular parallelepiped of the volume dxdydz from a uniform and isotropic infinite plate such as Fig. 5.3 whose sides are parallel to the axes of coordinates and are of lengths dx, dy, and dz.

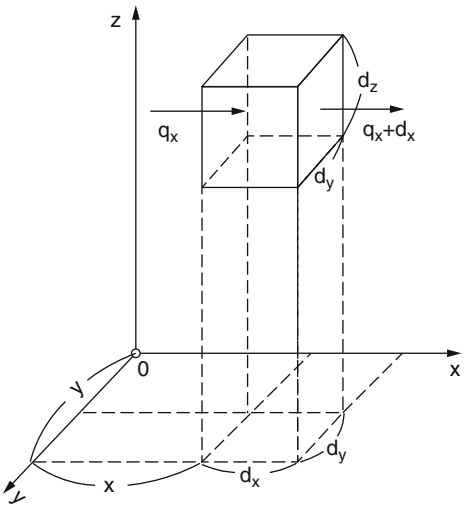


Fig. 5.3 Heat flux through elementary volume [3]

The fundamental differential equation of diffusion in an isotropic medium is derived from Eq. 5.1 as follows.

The heat amount flowing in through the left side, dydz, into the parallelepiped per unit time is $q_x$dydz, and the heat amount flowing out through the opposite side per unit time is $q_{x+\mathrm{d}x}$dydz.

Note that from now on we represent temperature with capital T instead of small t or C to keep everything in perspective and not to make any mistake with time t and we shall also convert τ to t for time.

However if $q_x > q_{x+\mathrm{d}x}$, then the elementary parallelepiped will be heated. But the difference between these flows according to the law of energy conservation is equal to the heat accumulated in this elementary parallelepiped, i.e.,

$$q_x \mathrm{dydz} - q_{x+\mathrm{d}x}\mathrm{dydz} = c\rho\frac{\partial T}{\partial t}\mathrm{dxdydz} \tag{5.11}$$

Note that the heat accumulated is calculated from the elementary relation $\nabla \dot{Q} = cM\nabla\Theta = c\rho V\nabla\Theta$, where $\nabla\Theta$ is a temperature increment in a body per unit time with the mass M and the volume V where c is representing specific heat and ρ is the density of this mass (i.e., $M = \rho V$).

The quantity of $q_{x+\mathrm{d}x}$ is an unknown function of x. If it is expanded in Taylor series (see Appendix A for Taylor series definition) and only the first two terms of the series are retained, it may be written as follows:

$$q_{x+\mathrm{d}x} \approx q_x + \frac{\partial q_x}{\partial x}\mathrm{d}x \tag{5.12}$$

Then substitution of Eq. 5.12 into 5.11 results

$$q_x\mathrm{dydz} - \left(q_x + \frac{\partial q_x}{\partial x}\mathrm{d}x\right)\mathrm{dydz} = c\rho\frac{\partial T}{\partial t}\mathrm{dxdydz}$$

$$q_x\mathrm{dydz} - q_x\mathrm{dydz} - \frac{\partial q_x}{\partial x}\mathrm{dxdydz} = c\rho\frac{\partial T}{\partial t}\mathrm{dxdydz}$$

$$-\frac{\partial q_x}{\partial x}\mathrm{dxdyd} = c\rho\frac{\partial T}{\partial t}\mathrm{dxdydz}$$

Using the heat conduction Eq. 5.7 in the form of $q_x = -k(\partial T/\partial x)$, we obtain

$$k\frac{\partial^2 T}{\partial x^2}c\rho\frac{\partial T}{\partial t}$$

or

$$\alpha\frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \tag{5.13}$$

$\alpha = k/c\rho$ is the thermal diffusivity, a material-specific quantity depending on the *thermal conductivity*, k, the *mass density*, ρ, and the *specific heat capacity*, c, and can be noted as c_p; from now on therefore α can be shown as $\alpha = k/c\rho$.

Equation 5.13 is a differential heat conduction equation for a one-dimensional heat flow. If heat propagates along the normal to the isothermal surfaces, then the vector **q** may be expanded in three components along the coordinate axes. The heat stored in the elementary volume of Fig. 5.3 will be equal to the sum of

$$-\left(\frac{\partial q_x}{\partial x}+\frac{\partial q_y}{\partial y}+\frac{\partial q_z}{dz}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

Then the differential heat conduction equation can be written as follows:

$$\alpha\left(\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{\partial^2 T}{\partial z^2}\right)=\alpha\nabla^2 T=\frac{\partial T}{\partial t} \tag{5.14}$$

where ∇^2 the Laplacian operator is the Cartesian coordinates and is presented as follows:

$$\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2} \quad \text{Cartesian coordinates}$$

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2} \quad \text{cylindrical coordinates}$$

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\theta}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \quad \text{spherical coordinates}$$

For details of derivation of the above relationship, refer to Appendix B.

This operator can be expressed in cylindrical and spherical as well and is shown in above as results the heat conduction equation can be expressed in those coordinates, respectively.

5.3.2.1 Diffusion Equation in Cylindrical Coordinate

Equation 5.14 can be expressed in cylindrical form by transformation of coordinates or by considering elements of volume of different shapes. Thus by putting

$$x = r\cos\theta$$
$$y = \sin\theta$$

or by considering an element of volume of a cylinder of sides dr, rdθ, and dz, we obtain the equation for diffusion in a cylinder as follows:

$$\frac{\partial T}{\partial t} = \frac{1}{r}\left\{\frac{\partial}{\partial r}\left(r\alpha\frac{\partial T}{\partial r}\right) + \frac{\partial}{\partial \theta}\left(\frac{\alpha}{r}\frac{\partial T}{\partial \theta}\right) + \frac{\partial}{\partial z}\left(r\alpha\frac{\partial T}{\partial z}\right)\right\} \tag{5.15}$$

in terms of the cylindrical coordinates r, θ, and z.

5.3.2.2 Diffusion Equation in Spherical Coordinate

The corresponding equation for a sphere in terms of spherical polar coordinates r, θ, and ϕ is obtained by writing

$$\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}$$

or by considering an element of volume of a sphere of sides dr, rdθ, and $r\sin\theta$dϕ. It is

$$\frac{\partial T}{\partial t}\frac{1}{r^2}\left\{\frac{\partial}{\partial r}\left(\alpha r^2\frac{\partial T}{\partial r}\right) + \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left(\alpha\sin\theta\frac{\partial T}{\partial \theta}\right) + \frac{\alpha}{\sin^2\theta}\frac{\partial^2 T}{\partial \phi^2}\right\}. \tag{5.16}$$

If the thermal conductivity is independent of temperature, then simplified forms of Eqs. 5.15 and 5.16 for pure radial diffusion, for example, in a long cylinder where end effects are negligible or in a spherically symmetrical system, can be expressed in terms of the nomenclature of vector analysis as follows:

$$\frac{\partial T}{\partial t} = \mathrm{div}(\alpha \mathrm{grad} T) = \alpha\nabla^2 T \tag{5.17}$$

For a one-dimensional symmetrical temperature filed, $\nabla^2 T$ is a function of one space coordinate.

5.3.3 *Boundary and Initial Conditions*

Before any solution can be presented for any governing diffusion equation, it is necessary to introduce additional physical information in the form of two conditions that are known as *initial* and *boundary* conditions which the temperature satisfies in the case of conduction problems. These are partly the direct expression of the results of experiment and partly the mathematical statement of hypotheses founded upon these results [5]. We assume that in the interior of the solid T is a continuous function of x, y, z, and t, that this holds also for the first differential coefficient with regard to t, and that this holds also for the first differential

coefficients with regard to x, y, and z. At the boundary of the solid, and at the instant at which the flow of heat is supposed to start, these assumptions are not made:

- *Initial condition* (I.C.) is a specification of the condition of the system at the start of the computation. The temperature throughout the body is supposed given arbitrary at the instant which we take as the origin of the time coordinate t. In a thermal and diffusion analysis problem, for example, the temperature of the melt at the time of laser dwelling with the target or some period of collapse time as an initial condition, similar argument can readily provide good starting value for other independent variables such as the velocity or the concentration of various chemicals. If this arbitrary function is continuous, we require finding a solution of our problem which shall, as t tends to zero, tend to the given value. In other words, if the initial temperature is given by

$$T = f(x, y, z),$$

  our solution of the equation

$$\frac{\partial T}{\partial t} = k\nabla^2 T$$

  must be such that

$$\lim_{t \to 0}(T) = f(x, y, z)$$

  at all points of the solid. If the initial distribution is discontinuous at points or surfaces, these discontinuities must disappear after a short period of time, and in this case our solution must converge to the value given by the initial temperature at all points where this distribution is continuous [6].
- The surface *boundary conditions* (B.C.s) usually arising in the mathematical theory of diffusion and heat transfer problem are described below. On the other hand, defining the specification of boundary conditions is normally more difficult. In some cases especially in the case of interaction of a high-power laser weapon with its target, very little is known about the actual thermal or dynamic of the target condition at the interface. However, there are basically three types of boundary conditions. The situation is simpler to analyze in the case of the thermal problem. The three types of boundary conditions for thermal problems are as follows:
  - At the outer boundary of the system, the temperature is specified $T = T_s$.
  - The heat flux is zero (insulating boundary which means *no flux across the surface*) $k\nabla T = 0$.
  - The heat flux is specified as q (heat exchange boundary, which means *prescribed flux across the surface*) $-k\nabla T = q_s$.

The book by Luikov (*Analytical Heat Diffusion Theory*) [3] offers different problems and associated solution for the above three types of boundary conditions, and readers should refer to that reference extensively to understand how to solve heat diffusion equation under these boundary conditions. Luikov also offers different methods for calculating the heat flow in the process of heating and cooling, where a body receives or releases a definite quantity of heat. His suggested method easily can be applied in the case of laser beam where the high-power laser beam is used as part of directed energy weapons.

5.3.4 Material Response

The laser heating of a material is mainly determined by the material's *absorptivity* (A) as function of laser wavelength. Accordingly the *reflectivity* (R) of materials also is dominated by the laser wavelength, and as a result both of these properties are main driving factors and criteria for material response to laser analysis. Absorptivity and reflectivity are part of optical properties of metals and in general the condition of their surfaces and temperature, i.e., the laser heating rate. The heating rate of the metal of consideration is mainly determined by the its absorptivity for a given laser wavelength—a quantity which, in return, is determined by the optical properties of the metal itself and of the metallic surface as well as the temperature range of the source, heating rate by the source, etc.

Given the above argument, the absorptivity (A) and reflectivity (R) are main driving factors for the choice of the most appropriate laser system as a mean for a directed energy weapon system. Surface hardening of the target of engagement also plays some part in this matter where the target lethality is under consideration. In the case of dealing with an incoming foe missile and destroying such threat in flight with a light beam such as laser, there are several possible approaches. One is to damage the missile's target seeker and prevent the missile from acquiring the target, while another case is to cause the warhead or rocket fuel to detonate prematurely. It is also possible to damage the flight controls and force the missile into an uncontrollable flight preplanned and path. The most common method is structurally weakening the missile body so that the missile breaks up in flight. Throughout these destruction methods, the ways in which missile materials react to laser irradiation are threefold:

1. Light coupling to the material—The optical reflectivity of the material determines what fraction of the energy is absorbed and thus converted to thermal and mechanical energy.
2. Propagation of thermal/mechanical effects—This characteristic determines the efficiency in which the heat or shock transmits through the material.
3. Induced effects of the propagation of thermal/mechanical energy—The resulting process occurs when high energy is deposited on a material, for instance, melting, vaporization, shock loading, crack propagation, and spalling.

5.3.4.1 Theory of Laser Interaction with Solids

Some of the first questions that should be asked when it comes to laser interaction with solids in particular if it is used as weapon to destroy a target are:

1. How much laser power is needed?
2. How long this power must be applied to dual target?
3. What side effects will be produced in addition to the heating process anticipated? Of course, the last question in the case of total target destruction is not important at this point.
   Finally the last but not the least is:
4. Are these requirements compatible with specifications of available laser system? Whether the ABL (airborne laser) or GBL (ground-based laser) or the system is orbiting beyond the Earth's atmosphere.

Naturally, ABL and GBL system requires certain other analyses, considerations, and effects such as thermal blooming, which causes the laser to defocus and disperse energy into the atmosphere. It can be more severe if there is fog, smoke, or dust in the air. As a result of the thermal blooming and atmospheric attenuation of the laser beam, the beam diverges and loses its energy that is supposed to be delivered to the target.

In many instances, the answers to the first three questions can be obtained by performing a few simple calculations based on classical heat transfer theory. Usually the results of these calculations will suggest an answer to the fourth question. The purpose of this discussion is finding some appropriate solution to the basic following heat equation under different boundary and initial conditions which are obtained for laser heating of solids (target) under a variety of conditions that pertain to the particular use case and applications [7]:

$$\nabla^2 T(x,y,z,t) - \frac{1}{\kappa}\frac{\partial T(x,y,z,zt)}{\partial t} = -\frac{A(x,y,z,t)}{K} \tag{5.18}$$

1. $A(x,y,z,t)$: Heat rate supplied to the target per unit time per unit volume
2. κ: Thermal diffusivity (units, cm^2/s)
3. K: Thermal conductivity (units, W/cm °C)
4. $T(x,y,z,t)$: Temperature (°C)

which are obtained for laser heating of solids under a variety of boundary, initial, and laser/target conditions that pertain to practical applications. The solution to Eq. D.18 can only be obtained in simple analytic form when one is prepared to make a variety of assumptions concerning the spatial and temporal dependence of the impressed laser heat source and the geometry of the sample that is being irradiated. As the description of these boundary conditions becomes more and more rigorous in terms of the actual spatial and temporal dependence of the heat source and the geometry of the workpiece, analytical solutions can no longer be

obtained, and the resulting expression for $T(x, y, z, t)$ can only be expressed numerically. Solutions to problems of this sort are of little use except in specialized studies and will not be discussed here. We will show that in many cases even quite crude approximations to the actual source and sample boundary calculations are capable of yielding predictions of $T(x, y, z, t)$ that correspond quite closely to actual temperature–time profiles in the solid. Where possible, these predictions have been generalized (i.e., expressed in reduced variable form), so that they may be applied to any material when thermal constants of that material are known [7].

One of the most important effects of high-power energy laser irradiation is the conversion of the optical energy in the beam into thermal energy into the material of target of interest. This is a classical heat transfer problem based on diffusion equation of 5.18, and we will summarize this thermal response using the initial and boundary condition defined above. The solution to conduction of heat transfer equation 5.18 in a three-dimensional solid is given in general by the solution to the following Cartesian form of heat transfer equation 5.19a:

$$\rho C \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(K \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(K \frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(K \frac{\partial T}{\partial z}\right) + A(x, y, z, t) \tag{5.19a}$$

or

$$\nabla^2 T - \frac{1}{\kappa} \frac{\partial T}{\partial t} = -\frac{A(x, y, z, t)}{K} \tag{5.19b}$$

where the thermal conductivity K, the density ρ, and the specific heat C are dependent both on temperature and position and heat is supplied by laser beam to target surface materials at the rate of $A(x, y, z, t)$ per unit time per unit volume [8, 9]. These thermal parameters and the temperature dependence on them make the Eq. 5.19a to be a nonlinear one, and in return solutions under different initial and boundary conditions become very difficult to obtain although numerical solutions are possible in a limited number of cases when the temperature dependence of $\kappa = (K/\rho C)$ [thermal diffusivity], K [thermal conductivity], ρ [material density of the target surface], and C [heat capacity or specific heat] is known. With simple assumption of those thermal properties of most materials that change greatly with temperature $T(x, y, z, t)$, they can often be assumed independent of temperature and can be assigned an average value for the temperature range of interest [7].

In order to support any experimental result of heat transfer data with theoretical calculations, we must have information on thermal parameters of the materials under consideration. See below:

- K: Thermal conductivity (units, W/cm °C)
- κ: Thermal diffusivity (units, cm^2/s)
- C_p: (or C) The heat capacity or specific heat at constant pressure (units, J/g °C) or ρC (units, J/cm^2 °C)

The average values of these parameters over the temperature range from 0 to $T\,^{\circ}\mathrm{C}$ can be designated as K_{avg} and κ_{avg} which mathematically can be written as follows [8]:

$$K_{\mathrm{avg}} = (1/T)\int_0^T K(T)\mathrm{d}T \tag{5.20}$$

$$\kappa_{\mathrm{avg}} = (1/T)\int_0^T \kappa(T)\mathrm{d}T \tag{5.21}$$

These integrals may be evaluated numerically when K and κ are not simple function of T. In the case of thermal properties varying with the temperature but independent of position, then Eq. 5.19a turns into the following form:

$$\rho C\frac{\partial T}{\partial t} = K\nabla^2 T + \frac{\partial K}{\partial T}\left\{\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2\right\} + A(x,y,z,t) \tag{5.22a}$$

Equation 5.22a clearly is a nonlinear case and under this condition may be reduced to a simpler form by introducing a new variable as follows [6]:

$$\Theta = \left(\frac{1}{K_0}\right)\int_0^T K\mathrm{d}T \tag{5.22b}$$

where K_0 is the value of K at $T = 0\,^{0}\mathrm{C}$. These, and the lower limit of integration, are merely introduced to give θ the dimensions of temperature and a definite value [6]. Note that Θ is essentially a potential whose gradient is proportional to the flux and then from Eq. 5.22b [6] follows that

$$\frac{\partial\Theta}{\partial t} = \frac{K}{K_0}\frac{\partial T}{\partial t},\quad \frac{\partial\Theta}{\partial x} = \frac{K}{K_0}\frac{\partial T}{\partial x},\quad \frac{\partial\Theta}{\partial y} = \frac{K}{K_0}\frac{\partial T}{\partial y},\quad \frac{\partial\Theta}{\partial z} = \frac{K}{K_0}\frac{\partial T}{\partial z}$$

This results in Eq. 5.19a to be reduced to the following equation:

$$\nabla^2\Theta - \frac{1}{K}\frac{\partial\Theta}{\partial t} = -\frac{A}{K_0} \tag{5.22c}$$

where, in Eq. 5.22c, A and $\kappa = (K/\rho c)$ are expressed as function of the new variable Θ; therefore in terms of this new variable, the heat conduction Eq. 5.19b is preserved with the condition that diffusivity κ now is a function of Θ [6]. In most cases the variation of κ with temperature is not important as K, so that, to a reasonable approximation, it can be considered to be a constant. For example, if a metallic surface is being near absolute zero, both K and c are approximately proportional to the absolute temperature. In such cases, if A is independent of T, Eq. 5.22c becomes of type Eq. 5.19b, and solutions for the case of constant conductivity may take over

immediately by replacing T by Θ, provided that the boundary conditions prescribed only T or $K\frac{\partial T}{\partial n}$; if they are of the form $\left(\frac{\partial T}{\partial n}\right) + hT = 0$ where h is a constant, this remark does not hold [5]. Note that $\partial/\partial n$ *represents differentiation along the outward-drawn normal to the surface*.

In steady-state cases, the situation is very important as well since Eq. 5.22c turns to Poisson's equation if A is constant and reduces to Laplace's equation if A is equal to 0. Then in these cases, finding solution to heat conduction problem is straightforward.

Another useful form may be obtained by introducing W, the heat content per unit mass of the material (measured from some arbitrary zero of temperature). In this case Eq. 5.19a reduces to the following equation:

$$\rho C\frac{\partial W}{\partial t} = \frac{\partial}{\partial x}\left(K\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(K\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(K\frac{\partial T}{\partial z}\right) + A(x,y,z,t) \tag{5.23}$$

Or, in terms of Θ defined by Eq. 5.22b, we have

$$\frac{\rho}{K_0}\frac{\partial W}{\partial t} = \nabla^2\Theta + \frac{A}{K_0} \tag{5.24}$$

where W is related to Θ in a known manner. The introduction of W has advantages in problems involving latent heat. We will further discuss the case of thermal properties varying with temperature and solving Eq. 5.22c for different boundary conditions imposed by the problem in hand, utilizing *Boltzmann's transformation* with constant diffusivity for the infinite composite solid scenario [5].

If the target or materials on the surface of the target are irradiated with a laser beam, the temperature in the vicinity of the focal spot on target will usually rise rapidly to within an order of magnitude.

More commonly, the approximation is made that Eq. 5.19b can be used with averaged values of the thermal constants over the temperature range of interest. Then Eq. 5.19b becomes

$$\nabla^2 T - \frac{1}{\kappa_{\text{avg}}}\frac{\partial T}{\partial t} = -\frac{A(x,y,z,t)}{K_{\text{avg}}} \tag{5.25}$$

In this case, solutions are possible for a number of cases in which thermal properties vary discontinuously (i.e., composite solids) or in those cases where a simple analytic expression is available for the spatial variation of K [7]. All these conditions are valued so long as we assume the solid is taken to be homogeneous and isotropic; then Eq. 5.19a reduces to Eq. 5.19b where again we have assumed that $\kappa = (K/\rho C)$ is the thermal diffusivity and holds. In the steady-state situation where $(\partial T/\partial t) = 0$, Eq. 5.19b reduces to

$$\nabla^2 T = \frac{A(x,y,z)}{K} \tag{5.26}$$

Note that in dynamic laser heating process where laser beam weapon interacts with a moving target, it may not be appropriate to use Eqs. 5.20 and 5.21 since these equations give equal weight to all temperatures in the range $0 \rightarrow T$ to determine a weighting factor for each $K(T)$ and $\kappa(T)$.

Both Eqs. 5.19b and 5.23 can be solved in a large number of cases using different methods such as separation of variables or utilizing Fourier or Laplace transform based on initial and boundary condition of the case in hand. Furthermore if no heat is applied to the surface of the material, the $A = 0$ and Eqs. 5.19b and 5.25 both reduce to

$$\nabla^2 T = \frac{1}{\kappa}\frac{\partial T}{\partial t} \quad \text{(transient-state case)} \tag{5.27}$$

$$\nabla^2 T = 0 \quad \text{(steady-state case)} \tag{5.28}$$

With appropriate boundary and initial condition applied in most cases whether heat source is present or absent, usually heat transfer problem (heat conduction) can be solved either by Eq. 5.24 or Eq. 5.26 by applying such boundary considering as heat flux transfer crosses the surface of the target or solid. In summary we can reduce the full conduction heat Eq. 5.18, that is, conduction with heat source or generation to very special cases. When the thermal conductivity K is constant, the first term of Eq. 5.18 becomes Laplacian of temperature T. The Laplacians of temperature in the three-principle coordinate system are listed in Table 5.1, while the general heat conduction equations with variable thermal conductivity, in three-principle coordinate systems, are listed in Table 5.2. The three special forms of the conduction Eq. 5.18 with constant thermal conductivity K are listed below:

1. **Laplace's equation**.
   This is for constant K, steady-state heat transfer so that the term $(\partial T/\partial t) = 0$, and no heat generation or $A = 0$, which is basically presented by Eq. 5.24 in above.

$$\nabla^2 T = 0 \tag{5.29}$$

   where $\nabla^2 T$ is a Laplacian of the temperature T.

Table 5.1 The Laplace of temperature in the three-principle coordinate

Coordinate system	$\nabla^2 T$
Rectangular	$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}$
Cylindrical	$\frac{\partial^2 T}{\partial r^2} + \frac{1}{r}\frac{\partial T}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2 T}{\partial z^2}$
Spherical	$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial T}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial T}{\partial \theta}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial^2 T}{\partial \phi^2}$

Table 5.2 Heat conduction equations with variable thermal conductivity in the three-principle coordinate systems

Coordinate system	$\nabla \cdot (K\nabla T) + A = \rho C \frac{\partial T}{\partial t}$
Rectangular	$\frac{\partial}{\partial x}\left(K \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(K \frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(K \frac{\partial T}{\partial z}\right) + A = \rho C \frac{\partial T}{\partial t}$
Cylindrical	$\frac{1}{r} \frac{\partial}{\partial r}\left(Kr \frac{\partial T}{\partial r}\right) + \frac{1}{r^2} \frac{\partial}{\partial \phi}\left(K \frac{\partial T}{\partial \phi}\right) + \frac{\partial}{\partial z}\left(K \frac{\partial T}{\partial z}\right) + A = \rho C \frac{\partial T}{\partial t}$
Spherical	$\frac{1}{r^2} \frac{\partial}{\partial r}\left(Kr^2 \frac{\partial T}{\partial r}\right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta}\left(K \sin\theta \frac{\partial T}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial}{\partial \phi}\left(K \frac{\partial T}{\partial \phi}\right) + A = \rho C \frac{\partial T}{\partial t}$

2. **Poisson's equation**.
   This is for constant K and steady-state heat transfer so that the term $(\partial T/\partial t) = 0$ with heat source being present; therefore $A \neq 0$.

$$\nabla^2 T + \frac{A(x, y, z)}{K} = 0$$

   or

$$\nabla^2 T + \frac{A}{K} = 0 \tag{5.30}$$

3. **Fourier's equation**.
   This is for constant K and no heat generation or $A = 0$ Which is basically the representation of Eq. 5.27 or transient state with no heat source:

$$\nabla^2 T = \frac{1}{\kappa} \frac{\partial T}{\partial t} \tag{5.31}$$

   The parameter κ is the thermal diffusivity, $\kappa = K/\rho C$.

In summary to find the solutions to various heat conduction problems, we need boundary conditions in space and time since both temperature T and heat generation term A are function of x, y, z and time t. In general, there are seven constants of integration. There is the first-order derivation with respect to the time variable and second-order derivatives with respect to each variable. The number of conditions for each independent variable is equal to the order of the highest derivative of that variable in the equation. Hence, one initial condition is required for all time-dependent problem; two boundary conditions are needed for each coordinate.

The spatial boundary conditions may be classified into three principal classes as we mentioned in Sect. 5.3.3 in above, and they can be summarized as follows [10]:

1. The first-kind or *Dirichlet* boundary conditions
2. The second-kind or *Neumann* boundary conditions
3. The third-kind or *Robin* boundary conditions

Each of these boundary conditions is described again as follows [10]:

1. First-Kind *(Dirichlet)* Boundary Conditions.
   Here, the temperatures are known at the boundaries.

$$T(\vec{x},t)|_{\text{surface}} = T_{\text{s}} \tag{5.32}$$

   An example of the first kind of boundary conditions for one-dimensional heat conduction is

$$T(x,t)|_{x=0} = T_0 \quad \text{and} \quad T(x,t)|_{x=L} = T_L$$

   An example of the first kind of boundary conditions for two-dimensional heat conduction is

$$T(x,y,t)|_{x=0} = T_0(y) \quad \text{and} \quad T(x,y,t)|_{x=L} = T_L(y)$$

   where T_0 and T_L are prescribed functions of y. If these functions are zero, these boundary conditions are called first-kind homogeneous boundary conditions.
2. Second-Kind (*Neumann*) Boundary Conditions.
   Here, the heat fluxes are known at the boundaries.

$$q_{\text{s}} = -K\frac{\partial T}{\partial x}\Big|_{\text{surface}} \quad \text{is known}. \tag{5.33}$$

   An example of the second kind of boundary conditions for one-dimensional heat conduction is
   $\frac{\partial T}{\partial x}\Big|_{x=0} = \frac{-q_1(y)}{K} = f_1(y)$ where f_1 is a prescribed function of y.
   If this function is zero, the boundary condition is called the second-kind homogeneous boundary condition.
3. Third-Kind (*Robin or Mixed*) Boundary Conditions.
   Here, the convection heat transfer coefficients are known at the boundaries.

$$q = h\Delta T = -K\frac{\partial T}{\partial \eta} \quad \text{is known}. \tag{5.34}$$

   An example of the third kind of boundary conditions for one-dimensional heat conduction is

$$h_1(T_\infty - T_{x=0}) = -K\frac{\partial T}{\partial x}\Big|_{x=0} \quad \text{or} \quad \left[-K\frac{\partial T}{\partial x}\Big|_{x=0} + h_1 T_{x=0}\right] = h_1 T_\infty = f_1$$

   where f_1 is a prescribed function of y.

Other boundary conditions include nonlinear-type boundary conditions. When there is radiation, phase change or transient heat transfer at the boundary conditions is nonlinear in nature.

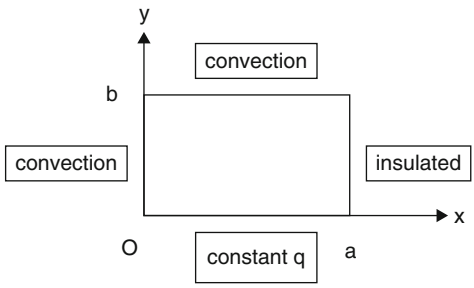


Fig. 5.4 Sketch for Example 5.1

Example 1 For a steady-state heat conduction problem with heat generation in a rectangular medium, write the governing equation and the mathematical representation of the boundary conditions. For $x = 0$, there is convection with heat transfer coefficient h_1. For $x = a$, the boundary is insulated. For $y = 0$, there is constant heat flux q. For $y = b$, there is convection with heat transfer coefficient h_2 (Fig. 5.4).

Solution The governing energy conservation (heat conduction) equation is given by

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{A}{K} = 0 \quad \text{for} \quad 0 \le x \le a \text{ and } 0 \le y \le b$$

The boundary conditions are

$$-K\frac{\partial T}{\partial x} + h_1 T = h_1 T_\infty \quad \text{at} \quad x = 0 \tag{5.35}$$

$$\frac{\partial T}{\partial x} = 0 \quad \text{at} \quad x = a \tag{5.36}$$

$$-K\frac{\partial T}{\partial y} = q \quad \text{at} \quad y = 0 \tag{5.37}$$

$$-K\frac{\partial T}{\partial x} + h_2 T = h_2 T_\infty \quad \text{at} \quad y = b \tag{5.38}$$

5.3.4.2 Laser Radiation, Effects on Solid Target

The fundamentals of which are given in this book, and particular Chaps. 5 and 6 are a systematic and in-depth study of the physical and chemical mechanism governing the interaction of laser radiation with solid targets, among them metals in different gaseous environments, and for a wide range of beam parameters.

The laser–solid interaction is crucial, particularly the amount of laser energy that is absorbed. It is shown that the dependence of the reflection coefficient of

Table 5.3 Characteristics of certain types of lasers

Laser type	Pulse duration (sec)	Pulse energy (J)	Power (W)	Maximum radiation flux density (W/cm^2)
CO_2	Continuous	–	10^3	Up to 10^7
Nd + glass	10^{-3}	10^4	10^7	Up to 10^7–10^{11}
CO_2	6×10^{-8}	3×10^2	5×10^{19}	10^{13}
Nd + glass	10^{-9}	3×10^2	3×10^{11}	10^{16}
Nd + glass	$(0.3) \times 10^{-11}$	10–20	10^{12}–10^{13}	10^{15}–10^{16}

absorbing media (metals) on the polarization of light may give rise to anisotropy of the absorption of powerful laser radiation in these media. This anisotropy is demonstrated by bending of the laser damage channel. When the polarization of light is linear, such bending occurs in a plane perpendicular to the plane of polarization, and the direction of bending is governed by an asymmetry of the distribution of the intensity in a cross section of the beam in this plane. The effect is weaker when the polarization is circular. The position of the bending plane is then governed by the nature of the asymmetry of the distribution of the intensity over the beam cross section.

The experiments have shown that the emission is essentially due to rapid heating of the surface spot which the laser beam strikes, and temperatures as high as 9000 K have been reached in some samples by using lasers with an output of 1 J [11]. Laser-induced electron and ion emission from metals have recently been observed by several authors [12, 13].

By focusing extremely high-power energy laser beam, radiation to its target production of light fluxes at the coupling interface with the target takes place.

The highest radiation power (see Table 5.3 below) has been produced using solid-state neodymium-doped glass lasers (wavelength $\lambda = 1.06$ microns [μ]) and gas lasers ($\lambda = 10.6$ μ). The specific features of laser radiation have led to the discovery of a number of new physical phenomena, the range of which is expanding rapidly as the power of lasers is increased.

The effect of high-power laser radiation in absorbing the solid target can be explained for different applications as follows:

Developed vaporization of metals. When laser radiation (e.g., pulses of a neodymium laser lasting several microseconds) with a radiation flux density of 10^6–10^8 watts per sq cm (W/cm^2) acts on metals, the metal in the zone of irradiation disintegrates, and a characteristic crater appears on the surface of the target. The bright luminosity of a plasma flare, which is a moving vapor heated and ionized by the laser radiation, is observed near the target. The reaction pressure of the vapor ejected from the surface of the metal imparts a recoil impulse Q to the target (Fig. 5.5).

Vaporization takes place from the surface of a thin layer of liquid metal heated to a temperature of several thousand degrees. The temperature of the layer is

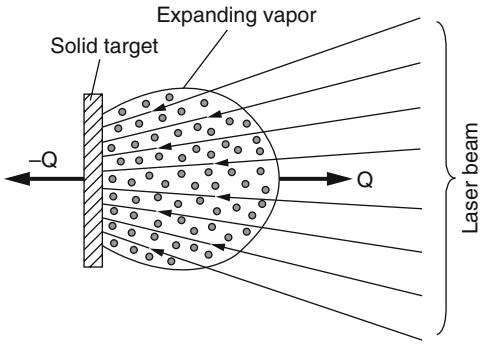


Fig. 5.5 Motion of vapor near the surface of a metal and transfer of momentum to a target from incident laser radiation: (**Q**) momentum vector of vaporized substance and (−**Q**) momentum imparted to solid target

determined by the equality of the absorbed energy and the losses to the cooling associated with vaporization. The role of thermal conduction in cooling the layer in the process is insignificant. In contrast to ordinary vaporization, this process is called developed vaporization.

The pressure in the layer is determined by the recoil force of the vapor and, when a gas-dynamic flow of vapor forms from the target, is one-half of the saturated vapor pressure at the surface temperature. Thus, the liquid layer is superheated, and its state is metastable. This makes it possible to study the conditions of maximum superheating of metals, under which rapid volumetric boiling-up of the liquid takes place. Upon heating to a temperature close to the critical temperature, an abrupt drop in electrical conductivity may take place in the liquid layer of the metal, and it may acquire the properties of a dielectric. In the process an abrupt drop in the light reflection coefficient is observed.

Irradiation of solid targets. As in the previous case, plasma is formed in the vapor flux from the vaporizing target upon irradiation of virtually all solid targets with millisecond pulses of laser radiation having a radiation flux density of the order of 10^7–10^9 W/cm^2. The plasma temperature is 10^4–10^5 K. This method may be used to produce a large quantity of dense, chemically pure low-temperature plasma to fill magnetic traps and for various industrial purposes. The vaporization of solid targets by laser radiation is used extensively in engineering.

When nanosecond laser pulses with a radiation flux density of 10^{12}–10^{14} W/cm^2 are focused on a solid target, the absorbing layer of the substance is heated so intensely that it immediately becomes plasma. In this case it is no longer possible to speak of vaporization of the target or of a phase interface. The energy of the laser radiation is used to heat the plasma and advance the disintegration and ionization front into the target. The plasma temperature is so high that multiply charged ions, in particular Ca^{16+}, are formed in it. Until recently, the formation of ions of such high multiplicity of ionization was observed only in the radiation of the solar corona. The formation of ions with a nearly stripped electron shell is also interesting from the standpoint of the possibility of conducting nuclear reactions known as ICF (inertial controlled fusion) using heavy nuclei in accelerators of multiply charged ions.

Laser spark (optical breakdown of a gas). When a laser beam with a radiation flux density of the order of 10^{11} W/cm^2 is focused in the air at atmospheric pressure, a bright burst of light is observed at the focal point of the lens, and a loud sound is heard. This phenomenon is called the laser spark. The duration of the burst exceeds the duration of the laser pulse (30 ns) by a factor of 10 or more. The formation of the laser spark may be represented as consisting of two stages: (1) the formation at the focal point of the lens of primary (seed) plasma, which ensures strong absorption of the laser radiation, and (2) the spread of the plasma along the beam in the area of the focal point. The mechanism of formation of seed plasma is analogous to the high-frequency breakdown of gases, hence the term "optical breakdown of a gas." For picosecond pulses of laser radiation ($I \sim 10^{13}$–10^{14} W/cm^2), the formation of seed plasma is also due to multiphoton ionization. The heating of the seed plasma by laser radiation and its spread along the beam (against the beam) are caused by several processes, one of which is the propagation of a strong shock wave from the seed plasma. The shock wave heats and ionizes the gas beyond the shock front, leading in turn to the absorption of the laser radiation—that is, to the maintenance of the shock wave itself and of the plasma along the beam (light detonation). In other directions the shock wave attenuates quickly.

Since the lifetime of the plasma formed by laser radiation greatly exceeds the duration of the laser pulse, at great distances from the focal point, the laser spark may be considered as a point explosion (the nearly instantaneous release of energy at a point). This explains, in particular, the high intensity of the sound. The laser spark has been studied for a number of gases at different pressures, under different conditions of focusing, and for various wavelengths of laser radiation, with pulses lasting 10^{-6}–10^{-11} s.

A laser spark may also be observed at much lower intensities if absorbing seed plasma is generated in advance at the focal point of the lens. For example, in air at atmospheric pressure, a laser spark is developed from electric-discharge seed plasma at a laser radiation intensity of approximately 10^7 W/cm^2; the laser radiation "captures" the electric-discharge plasma, and during the laser pulse, the luminosity spreads over the caustic surface of the lens. When the laser radiation is of relatively low intensity, the spread of the plasma is due to thermal conduction, as a result of which the rate of spread of the plasma is subsonic. This process is analogous to slow combustion, hence the expression "laser spark in the slow combustion mode."

Steady-state maintenance of a laser spark has been accomplished in various gases by means of a continuous CO_2 laser with a power of several hundred watts. The seed plasma was developed by a pulse CO_2 laser.

Thermonuclear fusion. Controlled thermonuclear fusion may be produced using laser radiation. For this purpose it is necessary to form extremely dense and hot plasma with a temperature of approximately 10^8 K (in the case of fusion of deuterium nuclei). For the energy liberation resulting from the thermonuclear reaction to exceed the energy added to the plasma during heating, the condition $n\tau \approx 10^{14}$ cm^{-3} s must be fulfilled, where n is the density of the plasma and τ is its lifetime or confinement time. For short laser pulses, this condition is satisfied at very high plasma densities. Here the pressure in the plasma is so great that

it is virtually impossible to contain it magnetically. The plasma that appears near the focal point disperses at a speed of the order of 10^8 cm/s. Therefore, τ is the time in which the dense plasmoid is unable to change its volume significantly (the inertial confinement time of the plasma). For thermonuclear fusion to occur, the length of the laser pulse t_1 obviously must not exceed τ. The minimum energy e of the laser pulse for a plasma density of $n = 5 \times 10^{22}\ \mathrm{cm}^{-3}$ (the density of liquid hydrogen), a confinement time of $\tau = 2 \times 10^{-9}$s, and a plasmoid with linear dimensions of 0.4 cm should be 6×10^5 J. Effective absorption of light by the plasma under conditions of inertial confinement and satisfaction of the condition $n\tau \approx 10^{14}$ occurs only for certain wavelengths λ and that is $\lambda_{cr} > \lambda > \left(\lambda_{cr}/\sqrt{40}\right)$, where $\lambda_{cr} \sim 1/\sqrt{n}$ is the critical wavelength for plasma with density n. When $n = 5 \times 10^{22}\ \mathrm{cm}^{-3}$, λ lies in the ultraviolet region of the spectrum, for which powerful lasers do not yet exist. At the same time, when $\lambda = 1\,\mu$ (a neodymium laser), even for $n = 10^{21}\ \mathrm{cm}^{-3}$, corresponding to λ_{cr}, a value of $e = 10^9$J for the minimum energy, which is difficult to realize, is obtained. The difficulty of feeding the energy of laser radiation in the visible and infrared bands into dense plasma is fundamental. Various ideas exist for surmounting this difficulty; one such idea that is of interest is the production of a superdense hot plasma as a result of adiabatic compression of a spherical deuterium target by the reaction pressure of plasma ejected from the surface of the target under the action of laser radiation.

High-temperature heating of plasma by laser radiation was accomplished in the first time by optical breakdown of the air. In 1966–1967, X-radiation from the plasma of a laser spark with a temperature of the order of $(1–3) \times 10^6$ K was recorded for a laser radiation flux density of the order of 10^{12}–10^{13} W/cm^2. In 1971 plasma with a temperature of 10^7 K (measured on the basis of X-radiation) was produced by irradiating a solid spherical hydrogen-containing target with laser radiation having a flux density of up to 10^{16} W/cm^2. A yield of 10^6 neutrons per pulse was observed in the process. These results, as well as the existing possibilities for increasing the energy and output of lasers, create the prospect of producing a controlled thermonuclear reaction using laser radiation [14–16].

Chemistry of resonance-excited molecules. A selective effect on the chemical bonds of molecules, making possible selective intervention in the chemical reactions of synthesis and dissociation and in the processes of catalysis, is possible under the action of monochromatic laser radiation. Many chemical reactions reduce to the scission of some chemical bonds in molecules and the formation of others. Interatomic bonds are responsible for the vibrational spectrum of a molecule. The frequencies of the spectral lines depend on the binding energy and mass of the atoms. A certain bond may be "built up" under the action of monochromatic laser radiation of the resonance frequency. Such a bond may easily be broken and replaced by another. Therefore, vibrationally excited molecules prove to be chemically more active (Fig. 5.6).

Molecules with differing isotopic compositions may be separated by means of laser radiation. This possibility is associated with the dependence of the vibrational frequency of the atoms comprising a molecule on the mass of the atoms. The monochromaticity and high power of laser radiation make possible selective

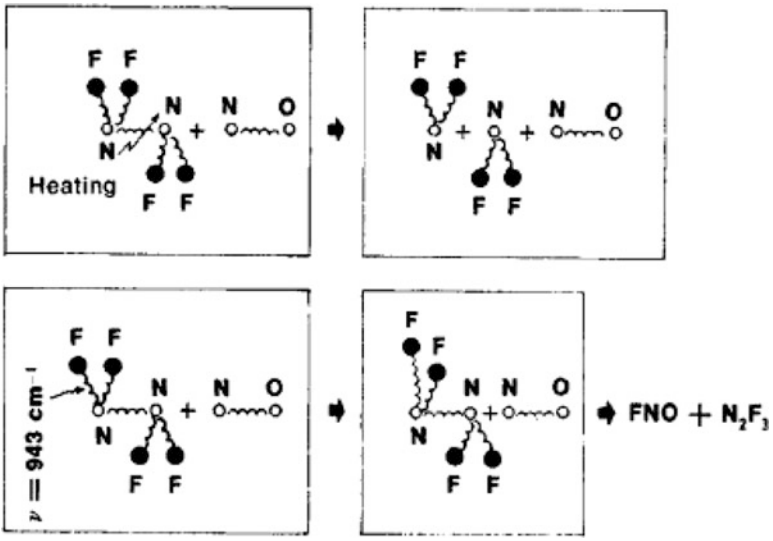


Fig. 5.6 Diagram of the reaction between tetrafluorohydrazine (N_2F_4) and nitric oxide (NO) upon heating (*top*) and upon resonance excitation of the N–F bond by laser radiation (*bottom*). The *wavy lines* represent chemical bonds

pumping of molecules of a specific isotopic composition to the predissociation level and the production of chemical compounds of monoisotopic composition or the isotope itself in the dissociation products. Since the number of dissociated molecules of a given isotopic composition is equal to the number of quanta absorbed, the effectiveness of this method may be high in comparison with other methods of isotope separation.

The effects mentioned above do not exhaust the physical phenomena caused by the action of laser radiation on the matter. Transparent dielectrics are destroyed under the action of laser radiation. When certain ferromagnetic films are irradiated, local changes in their magnetic state are observed. This effect may be used in developing high-speed switching devices and computer memory units. When laser radiation is focused within a liquid, the light-hydraulic effect, which makes possible the production of high pulse pressures in a liquid, occurs. Finally, for radiation flux densities of approximately 10^{18}–10^{19} W/cm^2, the acceleration of electrons to relativistic energies is possible. A number of new effects, such as the production of electron–positron pairs, are associated with this.

5.3.4.3 Absorption of Laser Radiation by Metals

The laser heating of materials is mainly determined by thermal parameters of the materials under consideration and particularly absorptivity of given metals at the respective laser wavelength. The metal absorptivity or, alternatively, the reflectivity

of the metal at a given laser wavelength plays major roles in laser interaction with metallic target. In this section we are trying to identify the key quantities of absorptivity/reflectivity as well as thermal conductivities to the optical properties of metals, the general condition of their surfaces, and temperature, i.e., laser heating rate through heat conduction equation under different boundary and initial conditions. We shall discuss the physical processes that occur during the interaction of high-power laser radiation with materials. Understanding of concept of these processes is important to have the knowledge and limitations of laser-based material processing. We pay attention and put emphases on metallic targets, but much of what we will discuss is applicable to other types of absorbing materials.

When laser duels with a target surface, part of it is absorbed and part is reflected. The energy that is absorbed begins to heat the surface. There are several regimes of parameters that should be considered, depending on the time scale, on the duration of engagement, and on the irradiance. The heating rate of metal sample or target surface is mainly determined by the target material absorptivity for a given wavelength—a quantity which, in turn, is determined by the optical properties of the metal itself and of the target surface, as well as by the temperature range, heating rate, etc. at the time of engagement. That is why the metal absorptivity, A, or alternatively the metal reflectivity, R, stands as the main criteria guiding the choice of the most appropriate laser system for destroying the metallic parts of the target or damages it enough that the incoming target is no longer a threat. For example, losses due to thermal conduction are small if the pulse duration is very short, but they can be important for longer pulses. Under some conditions, there can be important effects due to absorption of energy in the plasma formed by vaporized material above the target surfaces. We note that losses due to thermal reradiation from the target surfaces are usually insignificant.

The heating effects due to absorption of high-power beam can take place very rapidly. The surface temperature quickly rises to its melting point. Laser-induced melting is of interest because of target destruction during engagement. One often desires maximum melting under conditions where surface vaporization does not occur. Melting without vaporization is produced only within a fairly narrow range of laser parameters. If the laser irradiance is too high, the surface begins to vaporize before a significant depth of molten material is produced. In the case of industrial application of laser welding, this means that there is a maximum irradiance suitable for this purpose, but in our case where we are interested in target destruction, any damage to the target as the incoming threat falls as part of target lethality requirement that is mentioned in Sect. 5.3.4. Alternatively, for a given total energy in the laser pulse, it is often desirable to stretch the pulse length.

Melting of a material by laser radiation depends on heat flow in the material, which in turn depends on the thermal conductivity K. But on the other hand, the rate of thermal conductivity is not the only factor where the rate of change of temperature is depending on. The rate of change of temperature also depends on the specific heat c of the material at constant pressure. In fact Eq. 5.19b shows that the heating rate is inversely proportional to the specific heat per unit volume, which is equal to ρc, where ρ is the material density. The important factor for heat flow is

$K/\rho c$. This factor has the dimension of cm^2/s, which is per Eq. 5.19b is characteristic of the diffusion coefficient which is known as *thermal Diffusivity*.

The factor $K/\rho c$ is involved in all unsteady-state heat-flow processes, such as pulsed laser heating. The significant of this material property is that it determines how fast a material will accept and conduct thermal energy. Thus the higher thermal conductivity allows larger penetration of the fusion front with no thermal shock or cracking; on the other hand, lower thermal conductivity on the target material surface makes it harder for the laser to duel with it and limits the penetration of laser into the material, although low values of thermal diffusivity mean that the heat does not penetrate well into the material. But high value of thermal diffusivity can also allow rapid removal of heat from the surface, and this may cause reduction of melting amount. To compensate for these effects, one should vary the laser parameters for optimum effects for different materials. Table 5.4 lists the thermal diffusivity of several metals and alloys.

The depth of penetration of heat in time t is given approximately by the Eq. 5.39 below:

$$D = (4Kt)^{1/2} = \left(\frac{4Kt}{\rho c}\right)^{1/2} \tag{5.39}$$

where D is the depth of penetration of the heat and $k = K/\rho c$ is the thermal diffusivity. Typically for a metal with thermal diffusivity 0.25 cm^2/s, heat can

Table 5.4 Thermal diffusivity and thermal time constant [17]

Metal	Thermal diffusivity (cm^2/s)	Thermal time constants (ms)			
		0.01 cm thick	0.02 cm thick	0.05 cm thick	0.1 cm thick
Silver	1.70	0.015	0.059	0.368	1.47
Aluminum alloys					
Commercially pure	0.850	0.029	0.118	0.74	2.94
2024 alloy	0.706	0.035	0.142	0.89	3.54
A13 casting alloy	0.474	0.053	0.211	1.32	5.27
Copper alloys					
Electrolytic (99.95 %)	1.14	0.022	0.088	0.55	2.19
Cartridge brass	0.378	0.066	0.265	1.65	6.61
Phosphor bronze	0.213	0.117	0.470	2.93	11.74
Iron alloys					
Commercially pure	0.202	0.124	0.495	3.09	12.38
303 stainless steel	0.056	0.446	1.786	11.16	44.64
Carbon steel (1.22 C, 0.35 Mn)	0.119	0.210	0.840	5.25	21.01
Nickel alloys					
Commercially pure	0.220	0.114	0.454	2.84	11.36
Monel	0.055	0.455	1.818	11.36	45.46
Inconel	0.039	0.641	2.564	16.03	64.10

penetrate only about 3×10^{-4} cm during a pulse of 90-ns duration (typical of a Q-switched laser). During a pulse of 100-μs duration (typical of a normal pulse laser), heat can penetrate about 0.01 cm into the same metal.

The amount of light absorbed by the metallic surface is proportional to $1 - R$ which is the reflectivity. Quantitatively, the absorption A is the ration of the intensity absorbed by the metallic surface, I_a, to incident intensity, I, at a certain moment in the process of laser heating. Accordingly, the reflectivity, $R = 1 - A$, is the ratio between the reflected (specularly and/or diffused) intensity, I_r, and the incident intensity, I. The absorptivity of metals shows a general trend to increase when the incident radiation wavelength decreases from the infrared to the ultraviolet spectral range [8, 12]. At the CO_2 laser wavelength of 10.6 μm, where R is close to unity, $1 - R$ becomes small; as a result A is small. This means that only small fraction of the light incident on the surface is absorbed and is available for heating the engagement surface of the target. The difference in the value of R becomes important at long wavelengths. For copper or silver, for example, at 10.6 μm, $1 - R$ is about 0.02, whereas for steel it is about 0.05. Steel then absorbs about 2.5 times as much of the incident light as silver or copper. In practice, this means that steel surfaces are easier to engage with a CO_2 laser than are metals such as aluminum or copper.

The wavelength variation is also important. At shorter wavelengths, the factor $1 - R$ is much higher than at long infrared wavelengths. For example, the factor $1 - R$ for steel is about 0.35 at 1.06 μm, about seven times as great as its value at 10.6 μm. This means that, at least initially, seven times as much light is absorbed from a Nd:YAG laser than from a CO_2 laser for equal irradiance from the two lasers. In some cases it will be easier to carry out target destruction with a shorter-wavelength laser because of the increased coupling of light into the metallic surface of the target [17]. In general the shorter the wavelength, the better the coupling with the surface of the target.

Figure 5.7 shows some data on the reflectivity of a stainless steel surface struck by a 200-ns-duration pulse from a CO_2 TEA laser, which delivered an irradiance of 1.5×10^8 W/cm^2 to the target.

Note that one of the important factors for a battle management of some sort on ABL, SBL, or GBL platforms can possibly be to collect some information from its laser finder such as LADAR (laser detection and ranging) and target acquisition system that correct information using indirect determination of absorptivity A based upon measuring the reflectivity R. This method generally encounters its own several difficulties. First, we can note that the signal reflected by the target surface actually consists of a specular component as well as scattered (diffused) component—so that the value of total reflection coefficient R is $R = R_R + R_D$, where R_R and R_D are the coefficients of specular reflection and scattering, respectively.

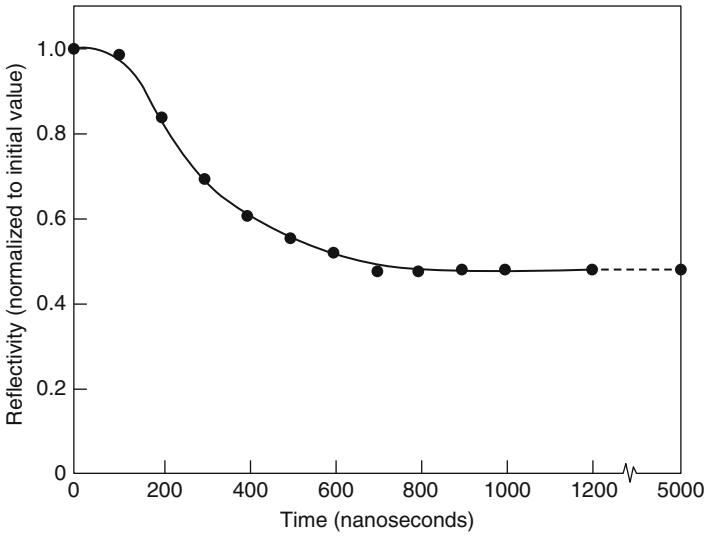


Fig. 5.7 Specular reflectivity at 10.6 μm as a function of time for a stainless steel surface struck by a CO_2 TEA laser pulse delivering 1.5×10^8 W/cm^2 in a pulse of 200 ns long [17]

(continued)

A FASOR used at the Starfire Optical Range for LIDAR and laser guide star experiments is tuned to the sodium D2a line and used to excite sodium atoms in the upper atmosphere.

LIDAR (light detection and ranging) is an optical remote sensing technology that measures properties of scattered light to find range and/or other information of a distant target. The prevalent method to determine distance to an object or surface is to use laser pulses. Like the similar radar technology, which uses radio waves, the range to an object is determined by measuring the time delay between transmission of a pulse and detection of the reflected signal. LIDAR technology has application in geomantic, archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing, and atmospheric physics [1]. Applications of LIDAR include ALSM (airborne laser swath mapping), laser altimetry, or LIDAR contour mapping. The acronym LADAR (laser detection and ranging) is often used in military contexts. The term "laser radar" is also in use even though LIDAR does not employ microwaves or radio waves, which is definitional to radar.

The ability of lasers to produce intense pulses of light energy leads to heating, melting, and vaporization of the target. The feature of laser that allows it to be used in directed energy weapons is, of course, its ability to deliver very high values of irradiance to a target surface. Irradiance can be defined as the incident laser power per unit area at the surface; it has units of W/cm^2. Only an electron beam can compare with a laser in this respect.

When laser radiation strikes a target surface, part of it is absorbed and part is reflected. The energy that is absorbed begins to heat the surface. There are several regimes of parameters that should be considered, depending on the time scale and on the irradiance. For example, losses due to thermal conduction are small if the pulse duration is very short, but they can be important for longer pulses. Under some conditions, there can be important effects due to absorption of energy in the plasma formed by vaporized material above the target surface. We note that losses due to thermal reradiation from the target surface are usually insignificant [17].

Since our interest on laser concentrates on its application as directed energy weapons, therefore the type of destruction effects on the incoming threat, whether it is melting or vaporization due to heat conduction from chosen laser, is not important. So as long as the threat or the incoming target gets destroyed or cannot deliver its assigned mission to its own preselected targets, then, we have achieved our mission of creating a DEW system.

In other applications of laser with the matter such as welding or cutting industries, the interest is on melting before we reach vaporization and plasma gets introduced. The heating effects due to absorption of high-power beams can occur very rapidly. The surface quickly rises to its melting temperature. Laser-induced melting is of interest because of welding applications. One often desires maximum melting under conditions where surface vaporization does not occur. Melting

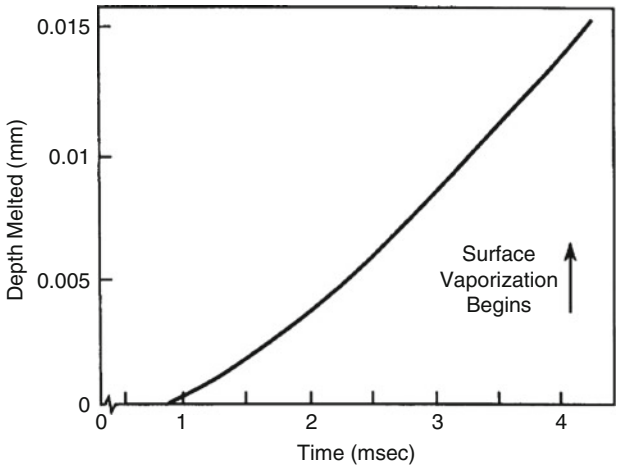


Fig. 5.8 Calculated depth melted in nick as a function of time for an absorbed laser irradiation of 10^5 W/cm^2

without vaporization is produced only within a fairly narrow range of laser parameters. If the laser irradiance is too high, the surface begins to vaporize before a significant depth of molten material is produced. This means that there is a maximum irradiance suitable for welding applications. Alternatively, for a given total energy in the laser pulse, it is often desirable to stretch the pulse length as we have described in above and presented in Eq. 5.39.

Effective melting and welding with lasers depend on the propagation of a fusion front through the sample during the time of the interaction, at the same time avoiding vaporization of the surface. Figure 5.8 shows the time dependence of the penetration of the molten front into a massive nickel sample for an absorbed irradiance of 10^5 W/cm^2. About 4 ms after the start of the pulse, the surface begins to vaporize. We note that the depth of penetration without surface vaporization is limited. To obtain greater depth, one can tailor the laser parameters to some extent. Generally, one lowers the irradiance and increases the pulse duration. The control is rather sensitive. One must make careful adjustments to achieve a balance between optimum penetration depth and avoidance of surface vaporization. (The results shown in Figs. 5.8 and 5.9 are calculated using an analog computer routine developed by M. I. Cohen [18].)

One is interested primarily in welding under conditions where surface vaporization does not occur. Melting without vaporization is produced only within a narrow range of laser parameters. If the laser irradiance is too high, the surface begins to vaporize before the fusion front penetrates deep into the material. This means that there is a maximum irradiance suitable for welding applications.

Alternatively, for a given total energy in the laser pulse, it is often desirable to stretch the pulse length to allow time for penetration of the fusion front through the workpiece. Figure 5.9 shows the depth of melting in stainless steel as a function of

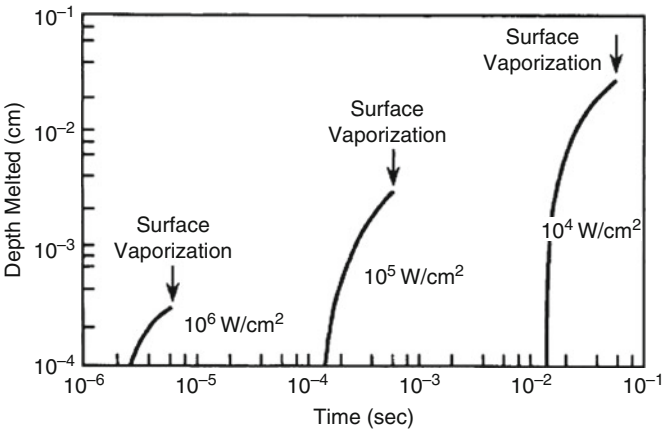


Fig. 5.9 Calculated depth melted in stainless steel for several different values of laser irradiance. The time at which surface vaporization begins is indicated for each curve [17]

time. Good fusion can be achieved over a range of pulse lengths if the laser energy is carefully controlled. For pulses shorter than 1 ms, surface vaporization is difficult to avoid [17].

One might think that one should use lasers with very high peak power in order to increase material removal. Paradoxically, it is not the highest laser powers that are optimal for material removal. The very high powers from a Q-switched laser vaporize a small amount of material and heat it to a high temperature. Early in the laser pulse, some material is vaporized from the surface. The vaporized material is slightly thermally ionized and absorbs some of the incident light. This heats the vapor more, producing more ionization and more absorption in a feedback process. Our earlier assumption that the laser light does not interact with the vaporized material is no longer valid if the irradiance becomes very high. Rather, the vaporized material does interact and absorb the incoming laser beam, so that the surface is shielded from the laser light [17]. Under some condition, most of the material may be removed as liquid. Figure 5.10 shows data on the fraction of the material ejected as liquid by a Nd–glass laser pulse with 30-kW power. Early in the pulse, most of the material was removed as vapor, but after a few hundred microseconds, about 90 % of the material removal occurred as liquid droplets.

Thus, new physical processes become important as the irradiance becomes very high. This is shown schematically in Fig. 5.11, which shows the depth vaporized as a function of time. The laser pulse shape is also shown for comparison. This is a typical pulse shape for Q-switched lasers. Early in the pulse, the surface starts to vaporize. Then the vaporized material heated and ionized by the laser forms a hot, opaque, ionized plasma, which absorbs essentially all of the incoming laser light. The flat portion of the curve represents the period when the surface is shielded by the plasma, so that vaporization ceases. Finally, late in the pulse, the plasma has expanded and become transparent again. Light can again reach the surface, and some additional material is vaporized. Because of these effects, the amounts of

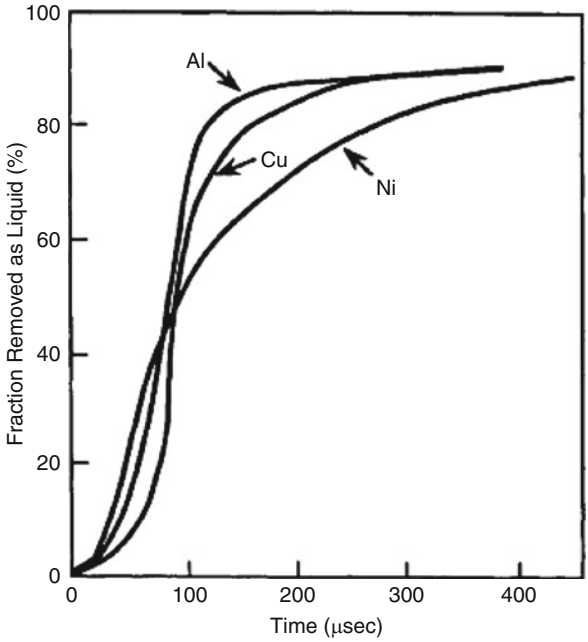


Fig. 5.10 Fractional of the material removed in liquid form. Results are shown as a function of time for several metals struck by a Nd–glass laser pulse (from M. K. Chun and K. Rose, J. Applied. Phys. 41, 614 (1970))

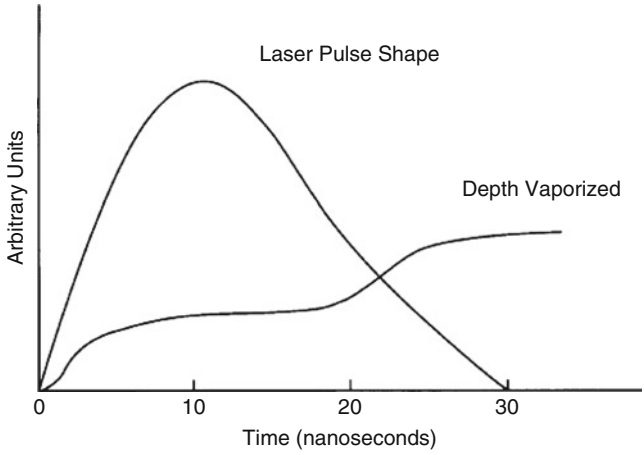


Fig. 5.11 Schematic representation of the depth vaporized in a metallic target as a function of time by a 30-ns-duration laser pulse with the indicated temporal profile. The effect of shielding of the target surface by blowoff material produced in the laser pulse is apparent [19]

material that can be removed by short-duration, high-power pulses, as from a Q-switched laser, are limited. Such lasers are not well suited for hole drilling or for cutting.

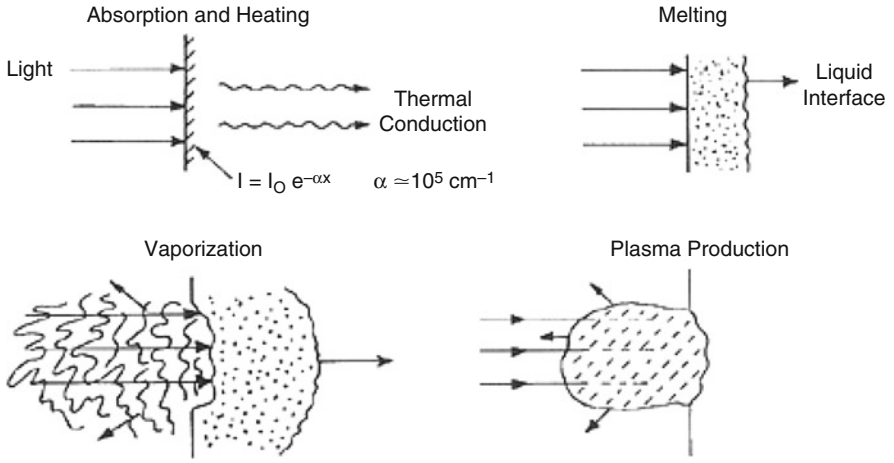


Fig. 5.12 Physical phenomena occurring when a high-power laser beam strikes an absorbing surface [17]

The shielding of the target by hot opaque plasma leads to a phenomenon called a laser-supported absorption (LSA) wave. The LSA wave is the plasma that is generated above the target surface and propagates backward along the beam path toward the laser. It is accompanied by a loud noise and a bright flash of light. Thus, the LSA wave makes an impressive demonstration, but while present it effectively shields the target surface and reduces the material removal. It can also drive a shock wave into the target.

We now try to explain one of the importance of physical phenomena that occur during the interaction of high-power laser radiation with the target surface, and we can summarize that again like we did in Sect. 5.3.4.2 in above. The physical phenomena are depicted in Fig. 5.12 below. If the top left portion of the figure indicates absorption of the incident laser light according to the exponential absorption law,

$$I(x) = I_0 e^{-\alpha x} \tag{5.40}$$

where $I(x)$ is the laser light intensity at depth x and I_0 is the incident laser light intensity, and α is the absorption coefficient. Based on this figure, the fraction of light that is reflected was neglected [17]. For metals, the absorption coefficient is of the order of 10^5 cm^{-1}. Thus, the energy is deposited in a layer about 10^{-5} cm thick. The light energy is transformed into heat energy essentially instantaneously, in a time less than 10^{-13} s. Thus the laser energy may be regarded as an instantaneous surface source of heat.

The heat energy then penetrates into the target by thermal conduction. When the surface reaches the melting temperature, a liquid interface propagates into the material, as indicated in the top portion of Fig. 5.10. With continued irradiation the material begins to vaporize, as indicated in the bottom left portion of Fig. 5.12,

Table 5.5 Approximate ranges of laser irradiance at which various processes dominate the laser–surface interaction [17]

Process	Range for visible and near-infrared laser (W/cm^2)	Range for far-infrared laser (W/cm^2)
Melting	-10^5	-10^5
Vaporization	10^6–1.5×10^8	10^6–2.5×10^7
Laser-supported absorption (LSA) wave	$>1.5 \times 10^8$	$>2.5 \times 10^7$
Plasma-collective effects	$\geq 10^{13}$	$>10^{13}$

and a hole begins to be drilled [17]. If the irradiance is high enough, absorption in the blowoff materials leads to hot opaque plasma. The plasma can grow back toward the laser as laser-supported absorption (LSA) wave. The LAS is defined in above. The plasma absorbs the light and shields the surface, as shown in the bottom right portion of Fig. 5.12. The ranges of laser irradiance for which individual processes dominate the interaction are given in Table 5.5. Values are stated for two wavelength regions: the visible and near-infrared region (around 0.5–1 μm) and the far-infrared region near 10 μm. The values in the table are approximate and will vary according to the exact parameters of the laser irradiation, such as pulse duration, target properties, and the like. At relatively low irradiance, melting is the main effect. At somewhat larger irradiance, vaporization becomes the most important effect. This is a conventional vaporization, with minimal interaction between the incident light and the vaporized material [17].

At still higher irradiance, LSA waves are kindled and dominate the physical processes, whereas vaporization is diminished. The thresholds for kindling the LSA waves are those appropriate to one specific case, namely, a titanium target with laser pulse duration in the microsecond regime. The threshold will vary as the circumstances change. But the numbers in Table 5.5 will serve to identify an order of magnitude at which certain types of interaction occur. The LSA wave dominates at a lower value of irradiance for far-infrared lasers than for visible and near-infrared lasers.

We summarize these phenomena in Fig. 5.13, which identifies various regimes of interaction and their potential applications. The figure defines these regimes in terms of irradiance and the duration of the interaction. The ordinate represents the pulse duration for a pulsed laser (or the time that the beam dwells on a spot for a continuous laser). Below the line marked “No Melting,” the surface is not heated to the melting point. In this region, one may have heat treating applications. In the region marked “Welding,” one obtains a reasonable depth of molten material, and welding applications are possible. Above the line marked “Surface Vaporization,” the surface begins to vaporize, and welding applications are less desirable. To the left of the welding region, the penetration of the fusion front is small because of the short interaction time. To the right of the welding region, the heat spreads over a broad area, and the desirable feature of localized heating is lost. Thus, welding operations usually require careful control to remain within this process window.

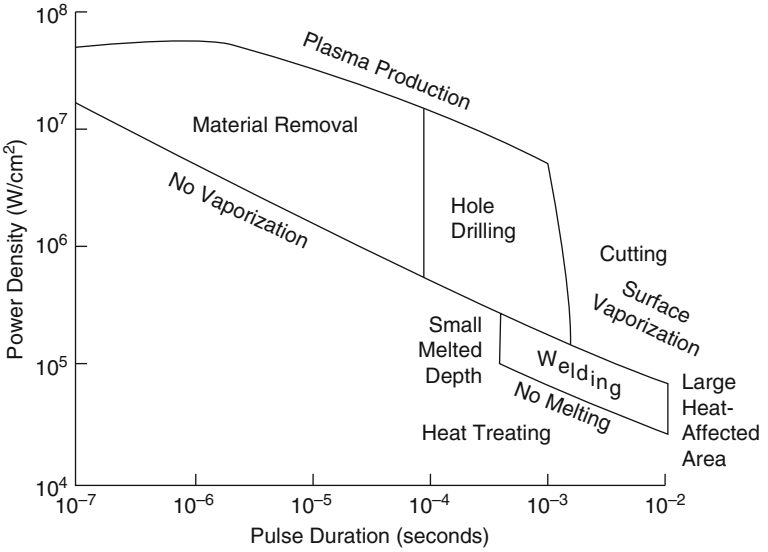


Fig. 5.13 Regimes of laser irradiance and interaction time for material processing application [17]

Similarly, the figure identifies regimes useful for cutting, hole drilling, and material removal for small amounts of material, such as vaporization of thin films, that is, trimming. Above the line marked "Plasma Production," the LSA wave develops. The only potential application identified in this region has been shock hardening [17].

The regions identified in Fig. 5.13 are not exact; they will vary with the target material, laser wavelength, and so forth. Still, they define regimes of laser parameters where certain applications are most likely to be productive. The engineer desiring to apply a laser in a specific material processing application must identify the process parameters suitable for that particular application [17].

5.3.4.4 Reflectivity at Normal Incident

Consider an electromagnetic wave in vacuum, with field components of the form [20]

$$E_y(\text{incident}) = E_y(\text{inc}) = B_z(\text{incident}) = Ae^{i(kx-\omega t)}$$

Let the wave be incident upon a medium of dielectric constant ε and permeability $\mu = 1$ that fills the half-space $x > 0$. Show that the reflectivity coefficient $r(\omega)$ as defined by $E(\text{refl}) = r(\omega)B(\text{inc})$ is given by

$$r(\omega) = \frac{n + ik - 1}{n + ik + 1}$$

where $n + ik \equiv \varepsilon^{1/2}$, with n and k real. Further we show that the reflectance is

$$R(\omega) = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2}$$

The reflected wave in vacuum may be written as

$$-E_y(\text{reflected}) = -E_y(\text{refl}) = B_z(\text{reflected}) = A' \mathrm{e}^{-i(kx+\omega t)}$$

where the sign of E_y has been reversed relative to B_z in order that the direction of energy flux (Poynting vector) be reversed in the reflected wave from that in the incident wave. For the transmitted wave in the dielectric medium, we find

$$\begin{aligned} E_y(\text{transmitted}) = E_y(\text{trans}) &= ck \frac{B_z(\text{transmitted})}{\varepsilon\omega} \\ &= \varepsilon^{-1/2} B_z(\text{transmitted}) A'' \mathrm{e}^{-(kx-\omega t)} \end{aligned}$$

by the use of the Maxwell's equation $\text{curl} H = \varepsilon \frac{\partial E}{\partial t}$ and the dispersion relation $\varepsilon\omega^2 = c^2k^2$ for electromagnetic waves.

The boundary conditions at the interface at $x = 0$ are that E_y should be continuous: $E_y(\text{inc}) + E_y(\text{trans}) or A + A' = A''$ or $A + A' = A''$. Also B_z should be continuous, so that $A + A' = \varepsilon^{1/2} A''$. We solve for the ratio A/A' to obtain $A + A' = \varepsilon^{1/2}(A - A')$, whence

$$\frac{A}{A'} = \frac{1 - \varepsilon^{1/2}}{1 + \varepsilon^{1/2}}$$

and

$$\boldsymbol{r} = \frac{E(\text{refl})}{E(\text{inc})} = -\frac{A}{A'} = \frac{\varepsilon^{1/2} - 1}{\varepsilon^{1/2} + 1} = \frac{n + (ik - 1)}{n + (ik + 1)}$$

The power reflectance is

$$R(\omega) = r^* r = \left(\frac{n - (ik-1)}{n - (ik-1)}\right)\left(\frac{n + (ik-1)}{n + (ik-1)}\right) = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2}$$

Some of the most interesting phenomena associated with lasers involve the effects produced when a high-energy power laser beam is absorbed at an opaque surface.

The most spectacular effects involve a change of phase of the absorbing material, for example, the luminous cloud of vaporized material blasted from a metallic surface and often accompanied by a shower of sparks [7].

For an opaque solid, the fraction of incident radiation absorbed is

$$\varepsilon = 1 - R_0$$

where ε is the emissivity and R_0 is the reflectivity at normal incidence. R_0 and ε can be calculated from measurements of optical constants or the complex refractive index. For a complex refractive index,

$$m = n - ik$$

then based on the above derivation, the reflectivity at normal incidence is

$$R_0 = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2}$$

The emissivity is then

$$\varepsilon = \frac{4n}{(n+1)^2 + k^2}$$

In general, n and k for metallic materials are functions of wavelength and temperature. The variation of n and k with wavelength and corresponding changes in ε for Ti at 300 K are shown in Fig. 5.14. It is apparent that both n and k are relatively slowly varying functions of λ over the range $0.4 < \lambda < 1.0$ μm and ε is large in this range. At longer wavelengths, n and k both increase rapidly with λ, and ε decreases to a small fraction of its value at shorter wavelength. In the infrared $\varepsilon \propto \lambda^{1/2}$ at constant temperature. Since $\varepsilon \propto r^{1/2}$, where r is the electrical resistivity, while r increases with temperature, $\varepsilon(\lambda)$ increases with temperature in the infrared [7]. The temperature dependence of $\varepsilon(\lambda)$ for $\lambda \leq 1$ μm is more complex; however,

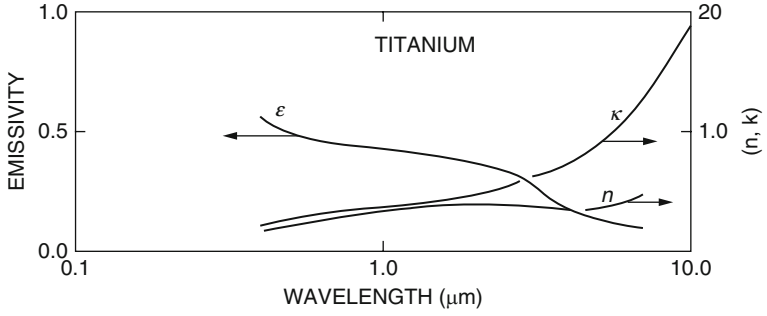


Fig. 5.14 Wavelength dependence of ε, n, and k for Ti at 300 K [21]

Table 5.6 Values of emissivity for various metals at laser wavelengths [21]

Metal	Emissivity[a]			
	Ar^+ (500 nm)	Ruby (700 nm)	Nd–YAG (1000 nm)	CO_2 (10 μm)
Aluminum	0.09	0.11	0.08	0.019
Copper	0.56	0.17	0.10	0.015
Gold	0.58	0.07	–	0.017
Iridium	0.36	0.30	0.22	–
Iron	0.68	0.64	–	0.035
Lead	0.38	0.35	0.16	0.045
Molybdenum	0.48	0.48	0.40	0.027
Nickel	0.40	0.32	0.26	0.03
Niobium	0.58	0.50	0.32	0.036
Platinum	0.21	0.15	0.11	0.036
Rhenium	0.47	0.44	0.28	–
Silver	0.05	0.04	0.04	0.014
Tantalum	0.65	0.50	0.18	0.044
Tin	0.20	0.18	0.19	0.034
Titanium	0.48	0.45	0.42	0.08
Tungsten	0.55	0.50	0.41	0.026
Zinc	–	–	0.16	0.027

[a]At 2 °C

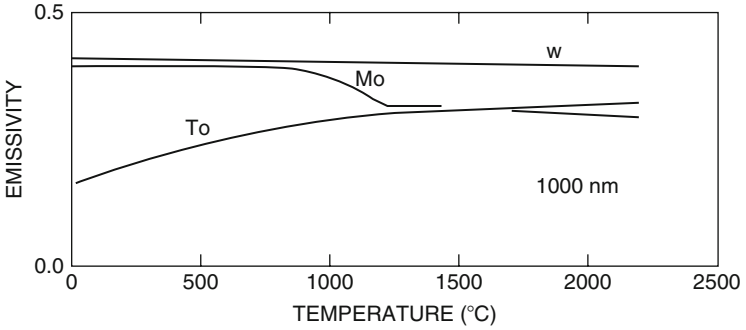


Fig. 5.15 Temperature dependence of emissivity for several metals at a wavelength of 1000 nm [21]

the net change in ε is smaller than what is observed in the infrared. In the visible region of the spectrum, ε often decreases slightly with increasing temperature [21].

Emissivities for metals at wavelengths characteristic of Ar^+, ruby, Nd–YAG, and CO_2 lasers are summarized in Table 5.6. The temperature dependence of ε for some metals at 1000 nm and 10.6 μm is shown in Figs. 5.15 and 5.16, respectively. Data on ε (1000 nm) were obtained from the observations of Barn. That for ε (10.6 μm) was calculated from the temperature-dependent emissivity from the expression given by Duley [8]:

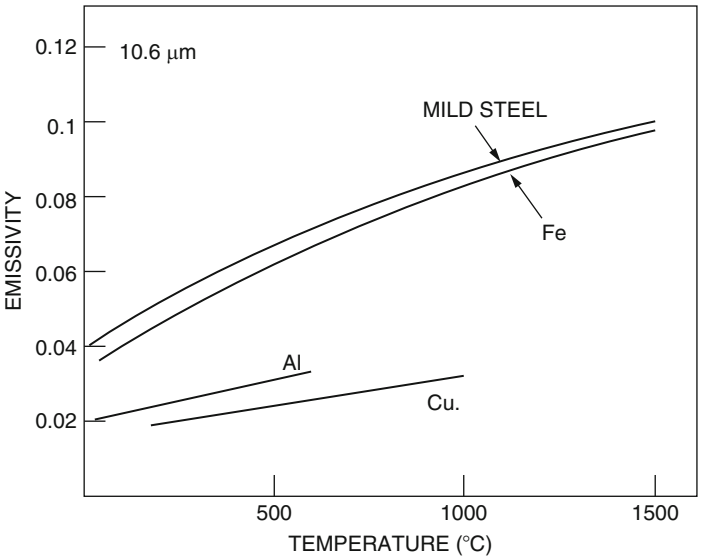


Fig. 5.16 Temperature dependence of emissivity at 10.6 μm for several metals [21]

$$\varepsilon_{10.6\mu m}(T) = 11.2\left[R_{20^\circ C}(1+\gamma T)\right]^{1/2} - 62.9\left[R_{20^\circ C}(1+\gamma T)\right] + 174\left[R_{20^\circ C}(1+\gamma T)\right]^{1/2} \tag{5.41}$$

where $R_{20^\circ C}$ is the resistivity at 20 °C and γ is the coefficient of resistivity change with temperature T °C.

Note: Equation 5.40 is valid for the metals heated in vacuum without a surface oxide layer. The presence of surface films will greatly increase $\varepsilon_{10.6\mu m}(T)$ [8].

Inspection of the data in Table 5.6 and Figs. 5.15 and 5.16 shows that the absorption of laser light by metallic surfaces at 20 °C is almost and an order of magnitude larger at visible wavelengths than at infrared wavelengths [21].

As we said in above, note the validation of these analyses takes place in the absence of any surface film. In most practical applications of laser heating, this assumption will not be valid because of oxide formation or the presence of surface contamination and possible other films (i.e., paints). When this is the case, values of ε in the infrared can be increased at visible wavelength. Thus, under practical conditions the difference between $\varepsilon_{10.6\mu m}$ and $\varepsilon_{visible}$ is unlikely to be as large as suggested by Duley [22] and data provided in Table 5.6. Duley et al. [23] have investigated the effect of oxidation on $\varepsilon_{10.6\mu m}$ for several metals heated in air. This follows work by Wieting and De Rosa [24] and Wieting and Schriempf [2] on the absorptance of stainless steel and Ti-6A1-4 V alloy at high temperatures in vacuum. Data obtained by Wieting and De Rosa on the absorptivity, ε, of type 304 stainless steel at 10.6 μm and presented by Duley [21] in his book are shown in different figures, and we suggest the reader to refer to his book.

The further study of interaction of high-power laser beam with materials and producing plasma at the target surface (see Fig. 5.12) will introduce to the research done by Sturmer and Von Allmen [3]. The time evolution of target shielding via plasma formation has been followed in details by them in 1978. They identify three separate absorption regimes when a high-intensity pulse of long duration is incident on a metallic target in air or other atmospheres. These are:

1. Strong reflection from the target
2. Absorption by plasma and target shielding
3. Dissipation of plasma and enhanced coupling to the target

As we discussed in the previous section to some degree, during step 1, which occurs during the initial stages of laser heating, the target has high reflectivity, and ε is small. The material evaporated from the focus may seed the gas in front of the target initiating plasma breakdown and lead to the formation of a laser-supported detonation (LSD) wave. This wave absorbs practically all the incident laser radiation (step 2) and shields the target. It dissipates by moving away from the target toward the focusing lens. This results in a reduction of plasma density followed by a decrease in opacity. The surface is then exposed to the last part of the laser pulse (step 3) which couples efficiently to the damaged target. Subsequent LSD wave ignition is suppressed by the low gas density left in front of the target following dissipation of the initial LSD wave.

While ε may be important in the initial stages of the heating of metallic targets with laser radiation, it is unimportant in many practical laser heating applications. The importance of ε is diminished when material removal has proceeded to the point where a cavity or keyhole has formed in the workpiece (target surface). In this case, the cavity acts as a blackbody absorber with ε effectively equal to unity. It has been shown [25] that control over the fitting of parts to be joined can also be effective in increasing ε. Under conditions in which laser radiation is absorbed in a keyhole, Steen and Eboo have shown that plasma absorption occurs within the keyhole yielding $\varepsilon = 1$.

The data contained in Table 5.6 can be used to obtain an estimate of the relative merit of Ar+, ruby, Nd–YAG, and CO_2 sources for laser heating. Heat transfer calculations show that the limiting temperature at the center of a Gaussian focal spot on a bulk target is

$$T = \frac{\varepsilon I_0 d \pi^{1/2}}{K} \tag{5.42}$$

where I_0 is the peak laser intensity (W/cm^2), d the Gaussian beam radius, and K the thermal conductivity. With optimum focusing $d \propto \lambda$, where λ is the laser wavelength, since $I_0 \propto P/\lambda^2$, where P is the laser power, one has

$$P \propto \frac{KT\lambda}{\varepsilon} \tag{5.43}$$

If we assume that useful thermal effects are produced only when $T = T_m$ or T_b, where T_m is the melting temperature while T_b is the boiling temperature, then this expression can be used to obtain an estimate of the relative difficulty of machining with different laser sources. A comparison of this sort is shown in Figs. 5.17 and 5.18 for the production of melting and boiling, respectively [21].

The correlations shown in these two figures are, of course, highly approximate because they assume room temperature values for ε and K can be used to high temperatures [22]. The uncertainty in these estimates can be reduced somewhat by taking $\varepsilon = 1$. A comparison calculated on this basis is shown in Fig. 5.19. If we compare powers required to melt or boil Ti from this figure, we see that $P(CO2)/P(Ar+) \sim 20{:}1$. Taking $\varepsilon < 1$ as shown in Fig. 5.14, the corresponding ratio are ~120.

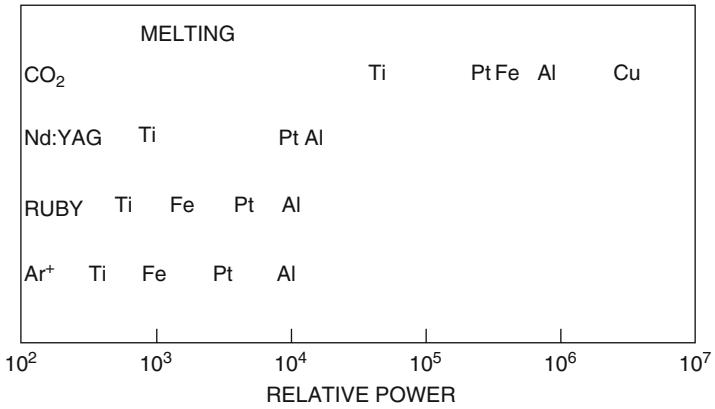


Fig. 5.17 Relative power to produce surface melting with various laser sources [21]

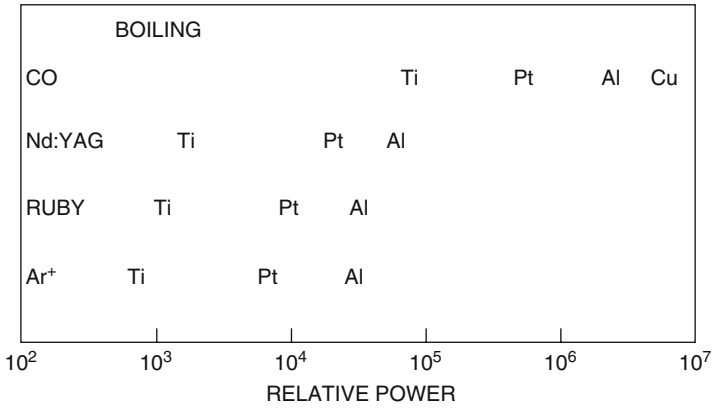


Fig. 5.18 Relative power to produce surface melting with various laser sources [21]

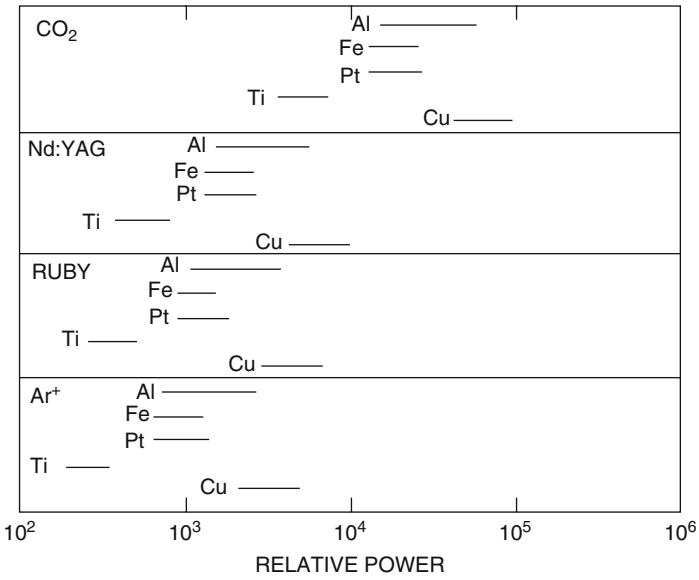


Fig. 5.19 Relative power to reach temperatures between T_m and T_b assuming $\varepsilon = 1$ [21]

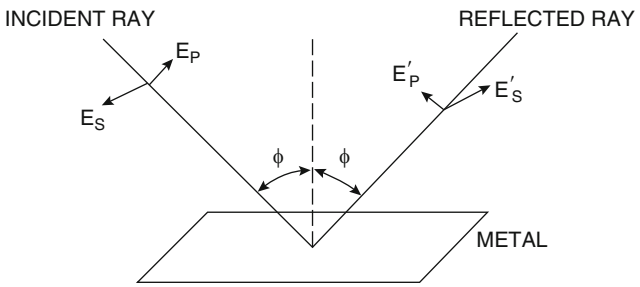


Fig. 5.20 Incident and reflected waves at a metallic surface [21]

Thus, while higher powers are required to initiate surface damage with infrared lasers than with visible lasers, when damage has occurred (i.e., $\varepsilon \sim 1$), much of the effect of the increased spot size of infrared lasers has disappeared [21].

At anything other than normal incidence, the reflection of laser radiation depends on polarization. This also well explained in Appendix F of this book. The geometry at some general angle of incidence ϕ is shown in Fig. 5.20. The reflectivity for the two polarization directions s and p, R_s and R_p, will in general be different. This means that the reflection coefficient for polarized laser light will be dependent on the orientation of the polarization vector relative to the metallic surface. An example of the angular dependence of R_s and R_p for Cu at 10.6 μm is shown in Fig. 5.21. It can be seen that R_s is high for all angles. However, R_p becomes very small at close to grazing incidence. Thus ε for incident light polarized

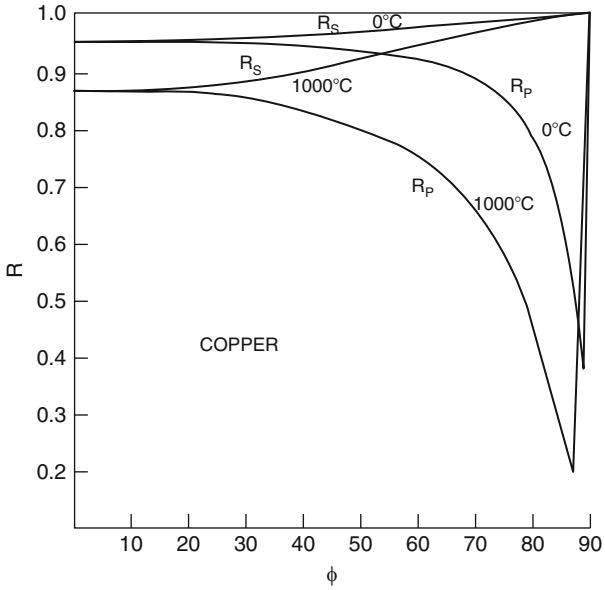


Fig. 5.21 Angular dependence of reflectivity for *s* and *p* polarizations with 10.6-μm radiation incident on Cu at 20 and 1000 °C [21]

perpendicular to the metallic surface is larger under these conditions, and enhanced coupling occurs. This has some important consequences in laser probing target where the efficiency of material removal depends on the relation between the polarization direction and the direction of translation of the metal substrate. This also can be demonstrated in the following section also known as *Fresnel absorption*.

5.3.4.5 Fresnel Absorption

Energy absorption by the workpiece from the laser can involve a direct process with the laser light incident on a surface as well as other indirect processes. There exists a simple electromagnetic model for absorption at a metallic surface that is widely used, especially in the context of keyhole modeling at the wavelength of a CO_2 laser. At that wavelength the assumptions of a simple model of electromagnetic interaction involving resistive dissipation are justifiable as a useful approximation, although at shorter wavelengths it becomes progressively more suspect. The model does not make allowance for surface impurities and must therefore be used with an understanding of its limitations.

This direct absorption process is usually referred to as *Fresnel abs*orption. A formula for the reflection coefficient $\Re$ that is frequently quoted [26], and which applies to circularly polarized light, is

$$\Re = \frac{1}{2}\left(\frac{1+(1-\varepsilon\cos\phi)^2}{1+(1+\varepsilon\cos\phi)^2}+\frac{\cos^2\phi+(\varepsilon-\cos\phi)^2}{\cos^2\phi+(\varepsilon-\cos\phi)^2}\right) \tag{5.44}$$

where ϕ is the angle of reflection that the light makes to the normal and ε is a material-dependent quantity defined by

$$\varepsilon^2 = \frac{2\varepsilon_2}{\varepsilon_1+\sqrt{\varepsilon_1^2+(\sigma_{st}/\omega\varepsilon_0)^2}} \tag{5.45}$$

where ε_0 is the permittivity of a vacuum, ε_1 and ε_2 are the real parts of the dielectric constants for the metal and the air or vapor through which the beam is being transmitted, and σ_{st} is the electrical conductance per unit depth of the workpiece. The value of ε_0 is 8.854×10^{12} F/m, and typical values for the other terms are approximately in unity for ε_1 and ε_2 and $5.0\times10^5\,\Omega$/m for σ_{st}. For a CO_2 laser, a wavelength of 10.6 μm gives a value for ω of $1.78\times10^{14}\ s^{-1}$, and so ε has a value of about 0.08. Figure 5.22 shows the graph of $\Re$ as a function of ϕ.

It will be seen that there is a strong dependence on the angle of incidence with a marked minimum close to $\phi = \frac{1}{2}\pi$, indicating that absorption is strongest at near-grazing incidence. For normal incidence, however, as much as 85 % of the incident

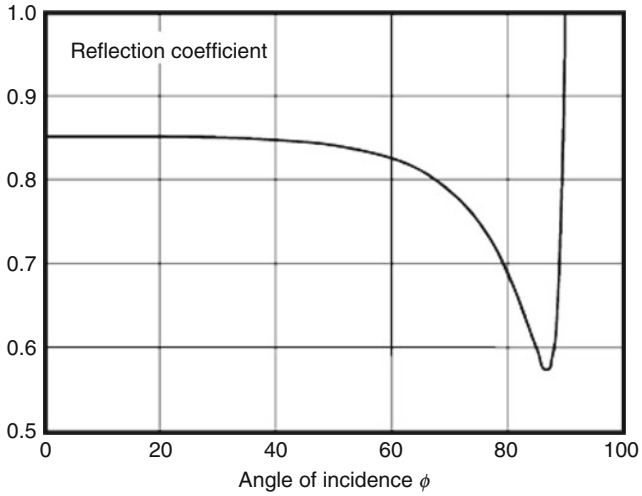


Fig. 5.22 The reflection coefficient $\Re$ as a function of the angle of the incident beam to the normal [27]

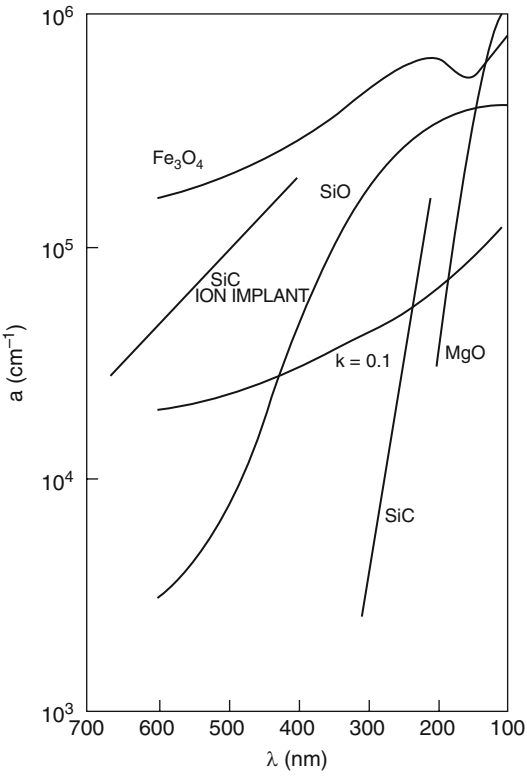


Fig. 5.23 Absorption coefficient α for several insulators at wavelengths between 100 and 700 nm [21]

power can be reflected. This figure can be very greatly modified by surface impurities or other additives introduced as part of the process.

Absorption of light by insulating materials is a strong function of wavelength. In the infrared, absorption arises from vibrational modes of the crystal lattice or organic solids by intermolecular vibration. Absorption coefficients $\alpha \sim 10^2 - 10^4$ cm^{-1} are typical within these bands. In the visible band, absorption may occur due to impurities (e.g., transition metal ions, crystal defect centers, etc.). Absorption can also occur due to discrete electronic transitions in molecular crystal (e.g., many organic solids) [22]. Absorption coefficients are typically $10^3 - 10^6 cm^{-1}$ within absorption bands. Figure 5.23 shows absorptions, α, for several refractory materials in the visible and ultraviolet α can be related to the transmission of a sheet of thickness t via

$$\frac{I}{I_0} \times 100 = \text{transmission in percent} = 100e^{-\alpha t}$$

or

$$\frac{I}{I_0} = \mathrm{e}^{-\alpha t} \tag{5.46}$$

where I_0 is the incident intensity and I is the transmitted intensity. A useful measure of the thickness required for significance of incident radiation is given by

$$L = \alpha^{-1}$$

where L is the attenuation length. A strong absorber has $\alpha = 10^6 \mathrm{cm}^{-1}$ and L cm, while a relatively weak absorber has $\alpha = 10^1 \mathrm{cm}^{-1}$ and $L = 10^{-1}$cm.

The relation between α and refractive index is given by Eq. 5.46 where k is an imaginary term in the complex refractive index

$$\alpha = \frac{4\pi k}{\lambda} \tag{5.47}$$

while λ is the wavelength of the incident light. In the visible region, nominally transparent materials will typically have $\alpha \sim 10^{-5}$ or $\alpha \sim 10\ \mathrm{cm}^{-1}$. The absorption due to $k = 0.1$ is shown in Fig. 5.23.

Note that in Fig. 5.23 the curve at $k = 0.1$ corresponds to the absorption that would be produced by a material with an imaginary refractive index equal to this value.

As we have mentioned, the optical properties of materials at UV wavelengths (i.e., typical CO_2 at $\lambda = 10.6$ μm) using the knowledge of electromagnetic radiation (see Appendix F) with condensed matter can be characterized in terms of a complex frequency-dependent dielectric constant $\varepsilon(w)$ [28]:

$$\varepsilon(w) = \varepsilon_1(w) + i\varepsilon_2(w) \tag{5.48}$$

where $\varepsilon_1(w)$ and $\varepsilon_2(w)$ are related to the complex refractive, m,as follows:

$$\varepsilon_1(w) = n^2 - k^2 \tag{5.49}$$

$$\varepsilon_2(w) = 2nk \tag{5.50}$$

with

$$m = n - ik \tag{5.51}$$

where n and k are both frequency dependent. Under an ideal vacuum circumstances, $n = 1$, and k is zero accordingly. The presence of matter causes both n and k to deviate from these values. With condensed matter, the density is many times larger than that of a gas. And deviations of n and k from vacuum values are correspondingly larger. It is not unusual to have $n, k \gg 1$ over a wide wavelength range in most solids.

Physically, the dependence of n on wavelength leads to dispersive effects in optical systems, whereas the absorption at a particular wavelength n is directly related to k. It can be shown that the real and imaginary terms either in the dielectric constant or in the refractive index are related through the Kramers–Kronig integrals [5]. For the dielectric constant $\varepsilon(w)$, these are [28]

$$\varepsilon_1(w) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{w^1 \varepsilon_2(w^1)}{(w^1)^2 - w^2} \mathrm{d}w^1 \tag{5.52}$$

$$\varepsilon_2(w) = \frac{-2w}{\pi} P \int_0^\infty \frac{\varepsilon_1 w^1 - 1}{(w^1)^2 - w^2} \mathrm{d}w^1 \tag{5.53}$$

where P is the principal part of the integral. These relations show that knowledge of either ε_1 or ε_2 over the frequency range $0 < w < \infty$ provides information on the value of the other at a specific frequency w. These relationships are often used to verify the consistency of experimental data for ε_1 and ε_2.

The absorption of light propagation through a medium is characterized by refractive indices and is given by Beer–Lambert's law [8]:

$$I(x) = I_0 \mathrm{e}^{-\alpha x} \tag{5.54}$$

where I_0 is the intensity at $x = 0$ and is the intensity after a distance $I(x)$. The attenuation coefficient

$$\delta^{-1} = \alpha = \frac{4\pi k}{\lambda} \tag{5.55}$$

is found to be directly proportional to the imaginary term in the refractive index. At UV wavelength, a transparent material would have $\alpha \leq 1\,\mathrm{cm}^{-1}$, whereas strong absorbers such as semiconductors or metals would have a $\alpha = (2-3) \times 10^6 \mathrm{cm}^{-1}$. The characteristic penetration depth (i.e., skin depth $\delta = \alpha^{-1}$) for radiation under these conditions is then α^{-1}. Note that Eq. 5.53 is valid only under conditions in which I_0 is much less than the intensity at which nonlinear effects may become significant. This can be shown in the following analysis and more details can be found in Schriempf report [9] as well. For the examples given, the penetration depth would be about 300–500 nm (metal) and $\geq$1 cm (transparent media). These values show that surface effects will dominate in the interaction of UV radiation with metals and many semiconductors. Surface roughness and composition are also important in determining the coupling of laser radiation to solids [9, 10, 29].

5.3.4.6 Optical Reflectivity

To consider the coupling of the laser energy to a material, we need first to know the optical reflectivity R and the transmissivity T for light incident on a surface which divides two semi-infinite media. The transmissivity plus the reflectivity equals unity at a *single surface*:

$$R + T = 1 \tag{5.56}$$

(See Appendix F for proof.) In most practical situations, we are dealing with more than one surface; typically, we have a slab of material with light impinging on one surface. Some light is reflected, and the rest is either absorbed or passes completely through the slab. In such a situation, we shall describe the net result of all the reflection, after multiple passes inside the slab and appropriate absorption has been accounted for, in terms of the reflectance R, the absorptance A, and the transmittance T:

$$\mathrm{R+A+T} = 1 \tag{5.57}$$

What we really are interested in from the point of view of material response is A, the absorptance of the material. In most materials of interest from the practical aim of using lasers to melt, weld, etc., T is zero, and

$$\mathrm{R} + \mathrm{A} = 1 \tag{5.58}$$

Schriempf [30] argues how to consider the relationship between R and R.

To understand reflectivity, we must use some general results from the theory of electromagnetic waves. Let us summarize these briefly at this point. The electric field of the electromagnetic wave, from the following Eq. 5.58, is

$$E = \mathrm{Re}\left[E_0 \mathrm{e}^{-2\pi kz/\lambda} \mathrm{e}^{i\omega(t-nz/c)}\right] \tag{5.59}$$

This equation is well defined in Appendix F of the book (volume 2). The relationships we need are those among the index of refraction n, the extinction coefficient k, and the material properties. These relationships can be derived by substituting Eq. 5.58 in the wave equation:

$$\frac{\partial^2 E}{\partial z^2} = \mu\varepsilon \frac{\partial^2 E}{\partial t^2} + \mu\sigma \frac{\partial E}{\partial t} \tag{5.60a}$$

where

E is the electric field of the radiation
Re stands for the real part of the complex quantity in brackets
E_0 is the maximum amplitude
k is the extension coefficient; in vacuum, $k = 0$
z is the direction in which the wave is propagating
λ is the wavelength
t is time
n is the index of refraction: in a vacuum, $n = 1$
c is the velocity of light in vacuum
σ is the electric conductivity or conductivity

μ is the magnetic permeability
ε is the dielectric function
ω is the angular frequency
This results in the following expression known as Eq. 5.60a:

$$\left(\frac{2\pi k}{\lambda}+\frac{i\omega n}{c}\right)^2=\mu\varepsilon\left(-\omega^2\right)+i\omega\mu\sigma \tag{5.60b}$$

Note that Schriempf [9] is using rationalized MKS units throughout. The material properties enter through μ, ε, and σ, which are the magnetic permeability, the dielectric function, and the electric conductivity of the medium. Using the usual equations between the field vectors as follows (see Appendix F and Eq. F.42 as well),

$$\begin{cases}\vec{D}=\varepsilon\vec{E}\\ \vec{B}=\mu\vec{H}\\ \vec{J}=\sigma\vec{E}\end{cases} \tag{5.61}$$

Plugging the following assumption in Eq. 5.60b, we get and the results final result into Eq. 5.60a along with some algebra work, we obtain Eq. 5.63:

$$\begin{cases}\varepsilon=K_{\mathrm{e}}\varepsilon_0\\ \mu=K_{\mathrm{m}}\mu_0\end{cases} \tag{5.62}$$

$$(k+in)^2=-K_{\mathrm{e}}K_{\mathrm{m}}\varepsilon_0\mu_0c^2+iK_{\mathrm{m}}\mu_0\sigma\frac{c^2}{\omega} \tag{5.63}$$

where
ε_0 is the electric permittivity of vacuum
μ_0 is the magnetic permeability of vacuum
K_{e} is the dielectric constant of metal
K_{m} is the magnetic permeability of metal
Finally, if we introduce $c^2=(\varepsilon_0\mu_0)^{-1}$ and some more algebra, we have

$$n-ik=\sqrt{K_{\mathrm{m}}}\sqrt{K_{\mathrm{e}}-i\frac{\sigma}{\varepsilon_0\omega}} \tag{5.64}$$

This equation relates the material parameters K_{m}, K_{e}, and σ, which in general may be complex, to the *index of refraction n* and *extinction coefficient k*. To describe the propagation of the light wave thus requires a knowledge of K_{e}, K_{m}, and a. Before we describe these, let us look at two more general properties of our propagating electromagnetic wave.

The first of these is absorption. If the medium is absorbing, the intensity will fall off to $1/e$ of its initial value in a distance δ, obtained by setting E^2 of Eq. 5.58 equal to $(1/e)E_{\max}^2$, or

$$\begin{cases} \dfrac{4\pi k\delta}{\lambda} = 1 \\ \delta = \dfrac{\lambda}{4\pi k} \end{cases} \tag{5.65}$$

From this equation we can see why k is called the *extinction coefficient*, for it determines the *skin depth* δ. Equation 5.64 is fairly general in that once k is known, δ can be calculated, providing that knowledge of the material of interest or target material properties is required to calculate k. Some information is discussed in Chap. 6 on how to measure these issues experimentally, and more can be found in a report by Joseph S. Accetta and David N. Loomis [31].

Schriempf [30] is deriving the second general property which we present here and is the expression for reflectivity, in terms of n and k again.

To do this, consider light impinging normally onto an ideal solid surface, as shown in Fig. 5.24. Here we have illustrated the incident E_i, reflected E_r, and transmitted E_t electric waves at a vacuum–material interface. For the present, we limit our discussion to the case of normal incidence. We now consider the boundary condition. We have for the electric field

$$E_i + E_r = E_t \tag{5.66}$$

For the magnetic field $\vec{B}$, we write

$$B_i - B_r = B_t \tag{5.67}$$

The minus sign is before B_r because $\vec{E} \times \vec{B}$ is positive in the direction of propagation of the wave. Now, the relationship between $\vec{B}$ and $\vec{E}$, or, since $\vec{B} = \mu\vec{H}$,

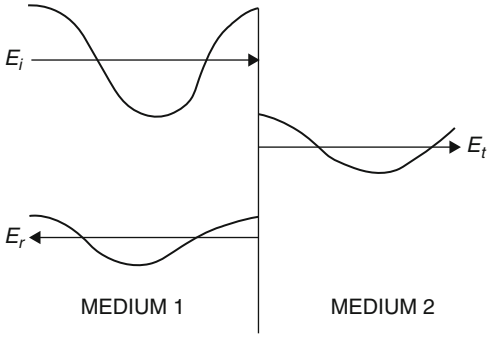


Fig. 5.24 Normal incident, transmitted, and reflected electric vectors at an interface

between $\vec{H}$ and $\vec{E}$, is required in order to proceed further. This follows directly from Maxwell's equations (see Appendix F):

$$\nabla \times \vec{E} = -\mu \frac{\partial \vec{H}}{\partial t} \tag{5.68}$$

$$\nabla \times \vec{H} = \sigma \vec{E} + \varepsilon \frac{\partial \vec{E}}{\partial t} \tag{5.69}$$

It is convenient to rewrite Eq. 5.58 and introduce $\omega\lambda = 2\pi c$, to have $\vec{E}$, explicitly in terms of ω instead of both ω and λ. Recall that $\vec{E}$ is a vector, and take it as being along the x-direction. Thus

$$E_x = E_0 \mathrm{e}^{i\omega t} \mathrm{e}^{-\frac{i\omega}{c} z(n-ik)} \tag{5.70}$$

Here we have dropped the "Re" notation and shall simply note that we always mean the real part when we write the wave in exponential form. We shall use unit vectors $\hat{\vec{x}}$ and $\hat{\vec{y}}$.

Now the curl expressions (see Appendix B on vector analysis) reduce to

$$\nabla \times \vec{E} = \hat{\vec{y}} \frac{\partial E_x}{\partial z}$$

which with Eq. 5.67 tells us that $\vec{H}$ has a y component:

$$\nabla \times \vec{E} = \hat{\vec{y}} \frac{\partial E_x}{\partial z} \tag{5.71}$$

Thus Eqs. 5.67 and 5.68 become

$$\frac{\partial E_x}{\partial z} = -\mu \frac{\partial H_y}{\partial t} \tag{5.72}$$

$$-\frac{\partial H_y}{\partial z} = \sigma E_x + \varepsilon \frac{\partial E_x}{\partial t} \tag{5.73}$$

and, of course, $E_y = E_z = H_x = H_z = 0$. Put the expression for E_x from Eq. 5.69 into 5.71 to find that

$$H_y = \frac{n - ik}{\mu c} E_0 e^{-\frac{i\omega}{c} z(n-ik)} e^{i\omega t}$$

This is the desired relationship:

$$H_y = \left(\frac{n - ik}{\mu c}\right) E_x \tag{5.74}$$

At this point we note in passing that Eq. 5.72 or 5.60a could be used to yield the relationship of n and k to μ, ε, and σ. If the reader is unfamiliar with these relationships, it is instructive to carry out the algebra. Returning to our consideration of the reflected electric and magnetic fields, we rewrite Eqs. 5.65 and 5.66 with the help of the relationship between $\vec{H}$ and $\vec{E}$, from Eq. 5.73:

$$E_\mathrm{i} + E_\mathrm{r} = E_\mathrm{t}$$

and

$$\mu_1 H_\mathrm{i} - \mu_1 H_\mathrm{r} = \mu_2 H_\mathrm{t}$$

becomes

$$E_\mathrm{i} - E_\mathrm{r} = \left(\frac{n_2 - ik_2}{n_2 - ik_1}\right) E_\mathrm{t}$$

Solve for $E_\mathrm{r}/E_\mathrm{i}$ by eliminating E_t:

$$\frac{E_\mathrm{r}}{E_\mathrm{i}} = \frac{n_1 - n_2 - i(k_1 - k_2)}{n_1 + n_2 - i(k_1 + k_2)}$$

Finally, the reflectivity R at the surface is

$$R = \left|\frac{E_\mathrm{r}}{E_\mathrm{i}}\right|^2 = \frac{(n_1 - n_2)^2 + (k_1 - k_2)^2}{(n_1 + n_2)^2 + (k_1 + k_2)^2} \tag{5.75}$$

Take medium 1 as a vacuum and drop the subscript 2. This gives, since in a vacuum $n_1 = 1$ and $k_1 = 0$,

$$R = \frac{(n_1 - 1)^2 + k^2}{(n_1 + 1)^2 + k^2} \tag{5.76}$$

Equation 5.75 is the second relationship we will find useful in discussing the coupling of optical relationship with metals. Note that it is derived for the special case of normal incidence and is applicable to a vacuum–material interface.

5.4 Effects Caused by Absorption of Laser Radiation at the Surface

Effects produced by a high-power laser beam focusing and absorbed by an opaque target surface are raising very interesting phenomena. The most spectacular effects involve a change of phase of the absorbing material such as the luminous cloud of vaporized material blasted from the metallic surface and often accompanied by a shower of sparks. The irradiance in the focal spot can lead to rapid local heating, intense evaporation, and degradation of the material. The most attractive feature of laser excitation is its capability to probe insulator within the focal spot and depositing heat energy into it. The most common mechanism of laser desorption is a thermally activated process induced by surface heating of the sample or surface of the target at the focal point. In this regime the amount of material transport across the surface is negligible (Fig. 5.25a). Laser heating of the solid surface and induced plume leads to the generation of different chemical species. Protonation and alkalinization reactions are often the source of the most characteristic species in the ion cloud [22]. Increasing the energy deposition into the sample, the surface temperature reaches a point where material transfer across the surface becomes significant (Fig. 5.25b).

The initial process in the conversion of high-power laser radiation to heat during irradiation involves the excitation of electrons to states of high energy. This is basically a simple conversion of optical energy of the beam into thermal energy in the material. This is the base of many laser applications including its weapon application as directed energy weapons. We shall summarize here this thermal response through its basic classical heat-flow problem and solving heat diffusion equation under different conditions.

Before we go further into this matter, we have to understand basic optical energy from atomic physics point of view and to understand how the principle of laser works. More correctly described a laser is a device for producing light that is almost

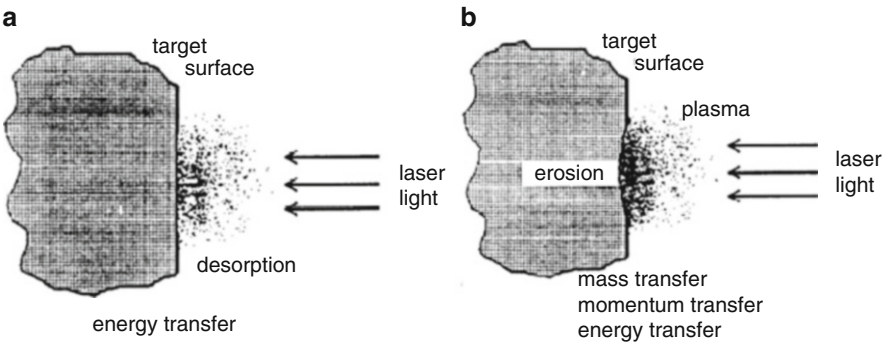


Fig. 5.25 Different regimes of laser–target interaction under the vacuum. In laser desorption (**a**) material transport across the surface is negligible. Laser volatilization (**b**) is characterized by considerable transport of mass, momentum, and energy and occasional plasma formation [32]

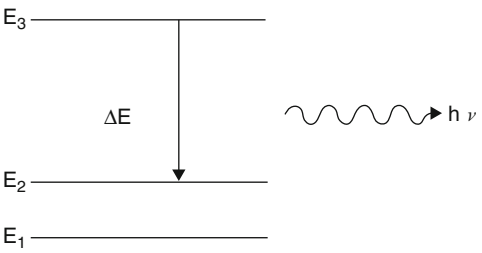


Fig. 5.26 Energy levels

totally coherent. It works in principle like this: an atom emits a photon of light when it decays from an excited energy state to a lower state; the difference in energy between the two states ΔE determines frequency v according to

$$\Delta E = hv \tag{5.77}$$

where h is Planck's constant. This is illustrated in Fig. 5.26. This is the case for any light source, whether laser, flame, incandescent body, etc. In the conventional light source, atoms emit photons in a random, sporadic manner and spontaneously decay to lower states when excited by heat or electric current. In a laser, on the other hand, the photons are emitted in the phase, and the electromagnetic radiation thus produced is, more or less, simply a propagating sinusoidal radiation field that can be described on a macroscopic level by Eq. 5.58 and further on by Maxwell's set of equations as described in above sections.

For this process to occur, vacant states have to be available to accept excited electrons. When the photon energy hv is small, as, for example, when 10.6-μm laser radiation is absorbed, only electrons within a narrow range hv near the Fermi energy, ε_F, can participate in absorption. At 0 K, the highest energy reached upon absorption is $\varepsilon_F + hv$.

At higher temperature, electrons occupy a range of states given by the Fermi–Dirac distribution. This reduces to a Boltzmann function for electron energies ε such that $\varepsilon - \varepsilon_F \gg kT$, where T is the metal temperature. Absorption of photons then populates those states with energy $\varepsilon_F + hv$. Since there are usually several electron volts, whereas $hv = 0.117$ eV for CO_2 laser photons, absorption of IR laser radiation then acts to distribute electrons among states close to those on the Fermi surface.

Fermi Energy

The Fermi energy is a concept in quantum mechanics usually referring to the energy of the highest occupied quantum state in a system of fermions at absolute zero temperature.

(continued)

The Fermi energy ε_F of a system of non-interacting fermions is the increase in the ground-state energy when exactly one particle is added to the system. It can also be interpreted as the maximum energy of individual fermions in this ground state.

Fermi Energy for Metal

As we defined above, the Fermi energy is the maximum energy that occupied an electron at 0 K. By the Pauli exclusion principle, we know that electrons will fill all available energy levels, and the top of that "Fermi sea" of electrons is called the Fermi energy or Fermi level. The conduction electron population for a metal is calculated by multiplying the density electron state $\rho(\varepsilon)$ times the Fermi–Dirac function $f_{FD}(\varepsilon)$. The number of conduction electrons permits volume per unit energy:

$$\frac{dn}{d\varepsilon} = \rho(\varepsilon) f_{FD}(\varepsilon) = \underbrace{\frac{8\sqrt{2}\pi m^{3/2}}{h^3}\sqrt{\varepsilon}}_{\text{Electron Density of state}} \underbrace{\frac{1}{e^{(\varepsilon-\varepsilon_F)/kT}+1}}_{\text{Fermi-Dirac distribution function}}$$

The total population of conduction electrons per unit volume can be obtained by integrating this expression:

$$n = \int_0^\infty \rho(\varepsilon) f_{FD}(\varepsilon) d\varepsilon = \frac{8\sqrt{2}\pi m^{3/2}}{h^3} \int_0^\infty \frac{\sqrt{\varepsilon}}{e^{(\varepsilon-\varepsilon_F)/kT}+1} d\varepsilon$$

At 0 K the top of the electron energy distribution is defined as ε_F so the integral becomes

$$\varepsilon_F = \left(\frac{(hc)^2}{8mc^2}\right)\left(\frac{3}{\pi}\right)^{2/3} n^{2/3}$$

For example, if we consider the element of Au having Fermi energy 5.53 eV, then the free-electron density is $n \simeq 0.5906466 \times 10^{29}$ electron/m^3.

This situation is different at excimer laser wavelength, since $h\nu$ is then comparable to or larger than the work function, φ, of many metals. When $h\nu > \varphi$, electrons may be directly excited from states near the Fermi surface to continuum states associated with the ejection of an electron from the metal. These electrons will originate from levels within the skin depth δ. Those electrons that are not ejected will dissipate their excess energy as heat within the skin depth. Photoelectrons, as they leave the surface with kinetic energy about $h\nu + kT - \varphi$, will *cool* the surface [28].

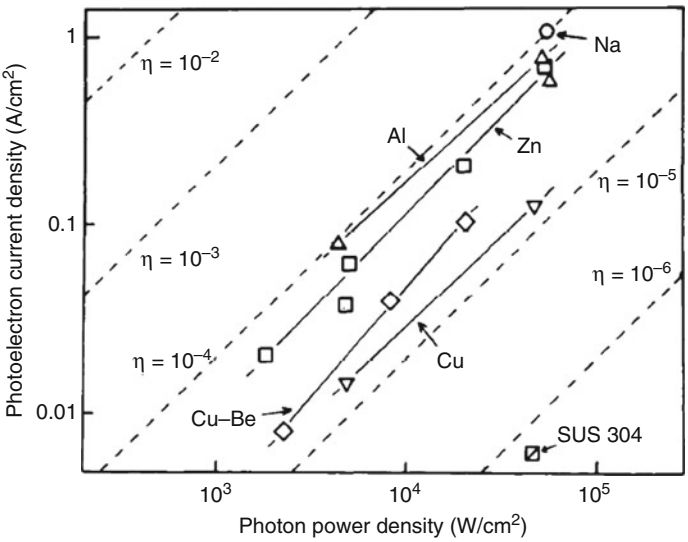


Fig. 5.27 Quantum efficiencies for photoelectron emission from various metals subjected to 248-nm KrF laser radiation [33]

Figure 5.27 shows a plot of photoelectron current density versus laser intensity for 248-nm KrF laser radiation incidents on several metals.

The transfer of energy from the electron gas to the metal to produce local heating and macroscopic thermal effects results from scattering of electrons by photons and defects.

In order to study pulsed laser heating and evaporation of solids, we constructed a one-dimensional model consisting of two parts: the first dealt with the heating and melting of the target and predicted the temperature, density, and flow velocity of the particles emerging on the liquid–vapor interface, whereas the second followed the expansion of the plume expelled from the surface. In the following sections, we present the framework of the calculations.

The distribution of this heat in response to a radiative source with defined spatial and temporal properties can then be calculated using the heat equation as follows:

$$\nabla^2 T(\vec{r},t) - \frac{1}{\kappa}\frac{\partial T(\vec{r},t)}{\partial t} = -\frac{A(\vec{r},t)}{K} \tag{5.78}$$

where

κ is the thermal diffusivity ($\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$) in MKS units

K is the thermal conductivity (W/cm/°C) in MKS units

In the above equation, $A(\vec{r},t)$ is the position-dependent rate of heat production per unit time per unit volume ($\mathrm{W/cm}^3$). Equation 5.77 assumes that K and κ are independent of temperature and do not vary across the metallic surface of the target.

As we discussed in Sect. 5.3.4.5 and Eq. 5.54, the deposition of heat under laser irradiation of opaque surface occurs over a depth (i.e., skin depth $\delta = \alpha^{-1}$) defined by the single-photon absorption coefficient $\alpha\,\mathrm{cm}^{-1}$. In metals, at optical frequencies this dimension is typically about 10^{-6} cm. Since α^{-1} is usually much smaller than the lateral spatial extent of focused excimer beam, the heat conduction equation (Eq. 5.77) can be linearized. Then in one-dimensional space, we can reduce it to the following form:

$$\frac{\partial^2 T(z,t)}{\partial z^2} - \frac{1}{\kappa}\frac{\partial T(z,t)}{\partial t} = -\frac{A(z,t)}{K} \tag{5.79}$$

where z is a coordinate extending from the sample surface into the material. Because $A(z,t)$ is a volume heat source, it must be evaluated over some incremental length Δz local at z. Assuming a homogeneous absorbing medium at the target surface, we can write

$$A(z,t) = (1-R)I_0(t)\alpha \mathrm{e}^{-\alpha z} \tag{5.80}$$

where R is the surface reflectivity and $I_0(t)$ is the time-dependent laser intensity incident at the surface. Then

$$\int_0^\infty A(z,t)\mathrm{d}z = (1-R)I_0(t) \tag{5.81}$$

Substitution of Eq. 5.80 into 5.78 results in the following equation:

$$\frac{\partial^2 T(z,t)}{\partial z^2} - \frac{1}{\kappa}\frac{\partial T(z,t)}{\partial t} = -\frac{1-R}{K} I_0(t)\alpha \mathrm{e}^{-\alpha z} \tag{5.82}$$

Equation 5.81 then describes the solution for the temperature profile produced inside a semi-infinite half-space exposed to a uniform surface heat irradiation by laser beam source of $(1-R)I_0(t)$ distributed as $\mathrm{e}^{-\alpha z} = \exp(-\alpha z)$ with depth known as the skin depth defined in the previous sections of this chapter.

The solution for the heat transfer Eq. 5.81 can be expressed in analytical form only when the system possesses certain symmetrical boundary conditions. When this is not the case, the heat transfer equation can be solved numerically. Detailed analysis of almost any practical laser heating problem requires that a numerical approach be adopted in particular for DEW application. However, as considerable physical insight into laser heating mechanisms can often be obtained from approximate solutions expressed in analytical form, available analytical solutions are summarized in this section.

These solutions, together with relevant boundary conditions, are shown presented by Duly [7]. While this does not represent a complete set of possible solutions, the examples shown have been chosen to be representative of boundary

conditions most often encountered in laser drilling. Further examples and discussion can be found elsewhere (Carslaw and Jaeger [6], Ready [19], and Duley [7]).

One form of these solutions is demonstrated in Sect. 5.4.1 below.

5.4.1 Heating Without Phase Change

In time of the duration of a laser pulse or irradiation, the electrons which absorb the photon will make many collisions, both among themselves and with lattice phonons. The energy absorbed by an electron will be distributed and passed onto the lattice. We can therefore regard the optical energy as being instantaneously turned into heat at the point at which the light was absorbed. The distribution occurs so rapidly on the time scale of Q-switched and normal laser pulses that we can regard a local equilibrium as established rapidly during the pulse. Therefore, the concept of temperature will be valid, and we are allowed to apply the usual equations for heat flow such as Eq. 5.81.

In the case that absorption coefficient α is relatively small and we are interested in the temperature at depth z of the order skin depth $\delta = 1/\alpha$, the solution is given by Eq. 5.82 below. The assumption is that the temporal pulse shape is flat (i.e., $A(t) = A_0 =$ constant for $t \geq 0$) and A_0 being the *incident intensity or beam flux* with dimension (W/cm^2 or J/cm^2 s). Under these conditions, in a material of thermal conductivity K and thermal diffusivity κ, the solution to the heat-flow Eq. 5.81 (the solution for this equation can be found using the Laplace transform method presented in Appendix E of this book) is given by

$$\begin{aligned} T(z,t) = {} & (2A_0/K)(\kappa t)^{1/2}\mathrm{ierfc}\left[z/2(\kappa t)^{1/2}\right] - (A_0/\alpha K)\mathrm{e}^{-\alpha z} \\ & + (A_0/2\alpha K)\exp(\alpha^2\kappa t - \alpha z)\mathrm{erfc}\left[\alpha(\kappa t)^2 - z/2(\kappa t)^{1/2}\right] \\ & + (A_0/2\alpha K)\exp(\alpha^2\kappa t + \alpha z)\mathrm{erfc}\left[\alpha(\kappa t)^2 + z/2(\kappa t)^{1/2}\right] \end{aligned} \tag{5.83}$$

In this equation, erfc and ierfc denote the complimentary error function and its integral that are well defined in Appendix E as well.

In the case that the optical absorption coefficient of the absorbing material will be larger for a typical metals where α is of the order of 10^5–10^6 cm^{-1}, then, the solution is proved in the following form:

$$T(z,t) = \left[2A_0(\kappa t)^{1/2}/K\right]\mathrm{ierfc}\left[z/2(\kappa t)^{1/2}\right] \tag{5.84}$$

For the case that large absorption coefficient is under consideration and varying temporal pulse shape with an infinite uniform spatial is extended, we may obtain the solution by applying Duhamel's theorem to Eq. 5.81 and present it as follows:

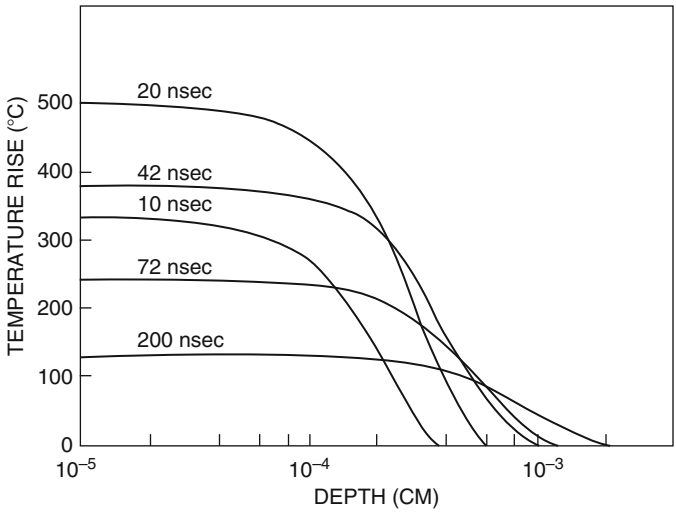


Fig. 5.28 Calculated temperature rise as a function of depth, with time as a parameter, caused by absorption of a Q-switched laser pulse in copper [21]

$$T(z,t) = \int_0^{\infty} \int_0^{t} \frac{A(\tau)}{A_0} \frac{\partial}{\partial t} \frac{\partial T'\left(z', t-\tau\right)}{\partial z'} \mathrm{d}z' \, \mathrm{d}\tau \tag{5.85}$$

where T' is the solution of the heat-flow equation for the case of a square pulse of absorbed flux density A_0.

Typical results for a calculated temperature rise as a function of depth, with time as a parameter, in a copper sample initially at 0 °C, are shown in Fig. 5.28 for the indicated laser pulse shape.

Further temperature profiles for a laser pulse with a typical shape discussion can be found in Ready [19].

5.4.2 *Heating with Phase Change*

When the surface temperature reaches the melting temperature of T_m, a melt region is formed adjacent to the surface. In the absence of perturbations, this molten material will propagate into the substrate at a speed υ_m that is given as following equation [34]:

$$\upsilon_\mathrm{m} = \frac{\varepsilon I_0}{\lambda_\mathrm{m} + \rho C T_\mathrm{m}} \exp\left(-\frac{\upsilon \Delta}{\kappa}\right) \tag{5.86}$$

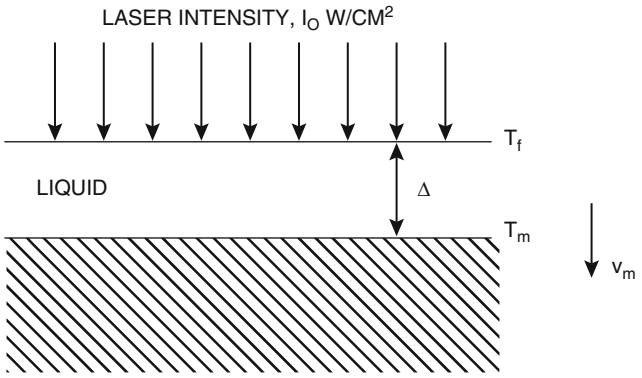


Fig. 5.29 Geometry of melt region on the surface of semi-infinite region heated uniformly over its surface

The geometry of this equation is depicted in Fig. 5.29 below, and λ_m is the latent heat of fusion with MKS dimension of (J km^3), C is the heat capacity holding the dimension of (J/g/°C), Δ is the melt thickness, and υ_m is in cm/s.

W. W. Duley [34] is an excellent discussion about dealing with this aspect of laser irradiation target and heating with the change of phase. If the liquid melt is not removed while the melting wave propagates into the solid, the temperature of the melt rises in response to the continuation of absorption at the melt–vapor interface. If the incident radiation is of sufficient intensity, then the temperature may rise to the boiling point, T_v, or higher. This is accompanied by the onset of a vaporization wave with a speed when $T = T_v$:

$$\upsilon_v = \frac{\varepsilon I_0}{\lambda_v + \rho C T_v} \tag{5.87}$$

where λ_v is the latent heat of vaporization and all quantities have the same units as in the case for υ_m. As I_0 increases, υ_v increases accordingly until it approaches the speed of sound υ_s in the material. When $\upsilon \to \upsilon_s$, then the equation for υ is written as follows:

$$\upsilon = \upsilon_s \exp\left(\frac{-\lambda Z}{\rho N_A k_B T_v}\right) \tag{5.88}$$

where

Z is the atomic number of material

N_A is the Avogadro number

k_B is the Boltzmann constant

Since υ is not dependent on I_0 in this regime, the speed of the vaporization rate saturates at high flux levels [34]. For most metals this saturation occurs when

$v \sim 10^5 - 10^6\ \mathrm{cm\,s^{-1}}$ and $I_0 \geq 10^8\ \mathrm{W\,cm^{-2}}$. Note that vaporization rates of this magnitude can be sustained only for short period of time, i.e., pulsed laser beam excitation [34]. Further discussion can be found in Carslaw and Jaeger [6], Ready [19], and Duley [7].

5.4.3 *Melt-Through of a Metal Plate*

Increasing the energy deposition into the sample target, it causes the surface temperature to reach a point where material transfer across the surface becomes significant (see Fig. 5.25b). Experimental observations by various researches show that target erosion appears in the form of craters. In theoretical terms it means that the energy balance equation has to be supplemented by the balance equations for mass and momentum phenomena. The flow, the heating, and the expansion of evaporated target material are governed by the equation of hydrodynamic. Solving the coupled partial differential equations of the conservation law provides insight into the factors determining crater depth, cloud extension, ion yield, relative sensitivity factors, and ion kinetic energy distributions. Calculated and measured values of these quantities show promising correlations for different kinds of lasers and different types of materials [35–37].

The depth of melting can be determined as a function of laser parameters (energy density and pulse duration in the case of pulsed laser and engagement or dual time for continuous laser beam).

Mathematical analysis and calculation are done by different researchers, and few that are recommended for further study are by Carslaw, H. S. and J. C. Jaeger [6], A. V. Luikov [4], and M. N. Ozisik [38].

The problem of heat conduction involving melting or solidification is complicated because the interface between the solid and liquid phase moves as latent heat is absorbed or librated at the surface, and under these types of regime, all the boundary conditions associated with such heat-flow problem more or less should be treated as *Lagrangian* rather than *Eulerian* types.

In fluid dynamics and finite deformation plasticity, the *Lagrangian* specification of the flow field is a way of looking at fluid motion where the observer follows an individual fluid parcel as it moves through space and time. Plotting the position of an individual parcel through time gives the path line of the parcel. This can be visualized as sitting in a boat and drifting down a river.

The *Eulerian* specification of the flow field is a way of looking at fluid motion that focuses on specific locations in the space through which the fluid flows as time passes. This can be visualized by sitting on the bank of a river and watching the water pass the fixed location.

(continued)

The Lagrangian and Eulerian specifications of the flow field are sometimes loosely denoted as the Lagrangian and Eulerian frame of reference. However, in general both the Lagrangian and Eulerian specification of the flow field can be applied in any observer's frame of reference and in any coordinate system used within the chosen frame of reference.

In such condition the location of the moving interface is not known a priori, and the thermal properties of solid and liquid are different. Ozisik [38] is handing solution of heat transfer flow equation at the moving interface including problems involving ablation. He also analyzes cases for a semi-infinite region with *variable surface heat flux* as well as *constant surface temperature* at $x = 0$. His approaches also deal with problems involving temperature-dependent thermal properties of target surface materials as well as time dependence.

5.4.3.1 Vaporization of a Target

Vaporization is very easy to produce with lasers. Vaporization by a high-power laser beam is a striking phenomenon. There is a shower of sparks characteristic of molten material expelled along with the vaporization. Induced plasma due to vaporization has its own implication on laser beam being absorbed by the surface of the target. This situation is to some degree has been discussed in Sect. 5.3.4.3 under the title of Absorption of Laser Radiation by Metals.

In the case utilizing laser as DEWs (directed energy weapons), vaporization of the target is not much of concern. By the time this phase is met, we are way beyond the target assigned mission in order to have any impact. Hopefully the incoming threat has lost all its momentum and path during the melting phase. But for those readers that are interested in studying this aspect of laser irradiation target, lots of literature and references are provided by different researchers including few that have been mentioned during the course of this chapter, and we recommend them to those types of readers.

References

1. McComb G (1997) Lasers, ray guns, & light cannons, project from the Wizard's workbench. McGraw-Hill, New York
2. Wieting TJ, Schriempf JT (1976) J Appl Phys 47:4009
3. Sturmer E, Von Allmen M (1978) J Appl Phys 49:5648
4. Luikov AV (1968) Analytical heat diffusion theory. Academic, New York
5. Wooten F (1972) Optical properties of solids. Academic, New York
6. Carslaw HS, Jaeger JC (1959) Chapter 1 Page 11 and 89 in Conduction of heat in solids, 2nd edn. Clarendon, Oxford

7. Duley WW (1976) CO^2 lasers: effects and applications. Academic, New York, NY
8. Birks JB (1970) Photophysics of aromatic molecules. Wiley-Interscience, London
9. Barbrino S, Grasso F, Guerriera G, Musumeci F, Scordino A, Triglia A (1982) Appl Phys A29:77
10. Roos A, Bergkvist M, Ribbing CG (1989) Appl Opt 28:1360
11. Honig RE (1963) Appl Phys Lett 3:8
12. Lichtman D, Ready JF (1963) Phys Rev Lett 10:342
13. Giovi F, Mackenzie LA, McKinney EJ (1963) Appl Phys Lett 3:25
14. Basov NG, Yu Gus'kov S, Danilova GV, Demchenko NN, Zmitrenko NV, Ya Karpov V, Mishchenko TV, Rozanov VB, Samarski AA (1985) Sov J Quantum Electron 15(6):852
15. Nuckolls JH (1980) In Laser Program Annual Report 1979, UCRL-50021-79, Vol 2, p. 2
16. Afanasiev YuV, Gamaly EG, Gus'kov SYu, Demchenko NN, Rozanov VB (1988) Laser and particle beam, 6, Paert 1:1
17. Ready JF (1997) Industrial applications of lasers, 2nd edn. Academic, New York
18. Cohen MI (1967) J Franklin Institute 283:271
19. Ready JF (1971) Effects of high-power laser radiation. Academic, New York, NY
20. Kittle C (1996) Introduction to solid state physics, 7th edn. John Wiley, New York, NY
21. Ready JF (1965) J Appl Phys 36:462
22. Vertes A, Juhasz P, Gijbels R, Fresenius Z (1989) Anal Chem 334:682
23. Duley WW, Semple DJ, Morency JP, Gravel M (1979) Opt Laser Technol 11:281
24. Wieting TJ, De Rosa JL (1979) J Appl Phys 50:1071
25. Shewell J (1977) Weld Des Fab, June, p 100
26. Stratton JA (1941) Electromagnetic theory. McGraw-Hill, New York, pp 500–511
27. Dowden JM (2001) The mathematics of thermal modeling: an introduction to the theory of laser material processing, 1st edn. Chapman and Hall, London
28. Duley WW (1996) UV lasers: effects and applications in materials science, 1st edn. Cambridge University Press, Cambridge
29. Kinsman G, Duley WW (1993) Appl Opt 32:7462
30. Schriempf JT (1974) Response of materials to laser radiation: a short course, NRL report 7728, July 10, 1974. Naval Research Laboratory, Washington, DC, 20375
31. Accetta JS, Loomis DN (2007) High energy laser (HEL) lethality data collection standards-revision A. Directed Energy Professional Society, Albuquerque, New Mexico. www.DEPS.org
32. Balaze L, Gijbels R, Vertes A (1991) Anal Chem 63:314
33. Kawmura Y, Toyoda K, Kawai M (1984) Appl Phys Lett 45:308
34. Duley WW (1983) Laser processing and analysis of materials. Plenum Press, New York
35. Vertes A, Juhasz P, De Wolf M, Gijbels R (1989) Int J Mass Spectrum Ion Processes 94:63
36. Vertes A, Juhasz P, De Wolf M, Gijbels R (1988) Scanning Microsc 2:1853
37. Vertes A, Juhasz P, De Wolf M, Gijbels R (1989) Adv A&s Spectrom 77:1638
38. Oziski MO (1968) Boundary value problems of heat conduction. Dover, New York

Chapter 6
High-Energy Laser Beam Weapons

In this section, we talk about beam weapons and their applications as directed energy weapons. The origin of laser technology dated back to a prediction made in 1916 by Albert Einstein where he suggested that an atom or molecule could be stimulated to emit light of a particular wavelength when light of that wavelength reached it, a phenomenon called stimulated emission. It had already been recognized that atoms and molecules emit and absorb light spontaneously, without outside intervention. In 1928, R. Ladenburg showed that Einstein's prediction was right. At that time, stimulated emission seemed to be a very rare occurrence that was inevitably overwhelmed by spontaneous emission. It would be many years before physicist learned how to create the right conditions to make practical use of stimulated emission in lasers, the physics of which we know today. The 1970s saw a series of breakthroughs that rekindled military interest in high-energy laser weaponry. These developments were centered in two areas, carbon dioxide and chemical lasers, the technology of which are known today. The CO_2 laser's potential for high-power output was recognized soon after it was first demonstrated by Patel although technology of gas dynamic laser was invented in 1967. Similar work was reported at about the same time by a Russian group which may have stimulated from American research works [1].

6.1 Introduction

Are laser and beam weapons purely defensive weapons that will protect us as if they were a colossal umbrella? Or are they dangerous and destabilizing new elements in an accelerated arms race? With the administration considering plans to spend billions of government and, as a result, tax payers' dollars on the research and development of laser and particle beam weapons, the time has arrived to consider seriously their implication.

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_6

The effectiveness of a defensive weapon system is measured by its ability to deny the attacking system succession accomplishing its mission. It takes much more than a powerful beam to make an effective weapon system. The beam must be aimed and focused through a generally uncooperative atmosphere or over long distances. It must concentrate high power on a small area long enough to do fatal damage, a requirement that typically means the beam must follow the target along its path. Most of the targets envisioned for beam weapons are fast enough that automatic tracking and identification are needed. In the case of systems for defense against nuclear attack, there should be a way to verify "kill" from far away.

There are several ways to look at the problems in destroying targets with beam weapons, and the job is complicated by the very complexities involved. The problems are best defined for laser weapons, and for that reason in this section, we concentrated primarily on lasers as directed energy weapons than other beam weapons such as particle beams or electromagnetic weapons which involve quite different physics and they have their own sections and are presented from their own sources. In laser beam weapons, there are several tasks involved, and some of might have common ground in respect to the other beam weapons, and they are as follows [2]:

1. Identify the target (which generally will not be sitting stationary by itself on a target range) and a vulnerable spot on it.
2. Track the target both until the weapon is ready to fire and while it is firing.
3. Point the weapon in the direction of vulnerable spot on the target.
4. Focus the beam so it has the desired intensity (generally the highest possible) at the target. In case laser beam as normal of incident as possible.
5. Compensate for atmospheric effects (i.e., thermal blooming) that otherwise would tend to make the beam wander off target or disperse the beam's energy.
6. Maintain focus of the beam on target during the attack.
7. Make sure that as much as possible the energy in the beam is deposited on the target, not deflected away from it. Verify that the target has been disabled.

These are all good wishing goals for a good DEW system, many of which are hard to achieve with the technology of lasers that we are aware of today. In order for DEW system to achieve such goals, military folks along with system developer's lump summing them under *target acquisition and fire contro*l. A process that provides detailed information about targets and locates them with sufficient accuracy to permit continued monitoring or target designation and engagement. This will include target acquisition for both direct and indirect fire weapons as well as for information operations that can be part of *artificial intelligence* (AI) system of DEW platform where the system is hosted and it is coupled with firing control of weapon as part of its command, control, communication, and intelligence ($_{C}^{3}{}_{I}$) overall system.

AI is nothing more than the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. Early target

discrimination is a very important task for an AI system coupled with DEW platform to shoot the right target since the total cost of ownership and operating cost for such DEW system in today's dollar is not cheap, and the shot from the firing system of DEW should not be wasted on the wrong target during its engagement. In this case, Surveillance and Target Acquisition is an important role assigned to AI units and/or DEW equipment. It involves watching an area to see what changes (surveillance) and then the acquisition of targets based on that information. So in an all beam control involves steering the beam and focusing it onto a target, in the process compensating for atmospheric distortions (i.e., thermal blooming) in case *ground-based laser* (GBL), *airborne-based laser* (ABL) where the beam goes through some layer of atmosphere. Off course any DEW in the form of *space-based laser* (SBL) does not fall into this category unless it is engaging in initial phase of target boosting from its ground lunching platform.

The emerging body of *Reconnaissance*, *Surveillance*, and *Targeting Acquisition* (RSTA) resources brings a powerful contribution to battle-space domination. Diverse RSTA operations occur simultaneously within the battle space—keyed to support a range of users from decision makers to "shooters." In addition to collecting information that develops situational awareness, RSTA assets contribute too many battle-space activities: intelligence preparation of the battle space, indications and warning, situation development, force protection, battle damage assessment, and targeting and collection queuing. Given this multidimensional capability, it is no longer desirable to relegate RSTA assets solely to the realm of intelligence collection management. The command and control of finite, high-value RSTA resources is ultimately the commander's responsibility.

Surveillance and Target Acquisition is a military role assigned to units and/or their equipment. It involves watching an area to see what changes (surveillance) and then the acquisition of targets based on that information.

Fire control in conjunction with an RSTA system is the aiming and firing of the weapon, a task which includes identifying and tracking targets and providing information which the beam control system can use to point the beam (AI) and triggering shots from the weapon, spotting vulnerable points on the target, and making sure that target is crippled or destroyed.

Beam control is unique to directed energy weapons, but fire control is a well-established military technology that is used in many kinds of weapon systems. Interestingly, the fire control systems used with many modern missiles and "smart" bombs incorporate low-power lasers to mark potential targets with a laser spot that can be detected by a sensor in the bomb or missile. Low-power lasers can also measure the distance to a target to aid artillery fire control system in pointing weapons. But although fire control equipment is used widely in conventional weapons and battlefield of today, directed energy weapons would demand more than current equipment and existing technology that can deliver. Indeed, many observers believe that beam and fire control are much more difficult problems than building a big laser. The difficulties arise because of the very demanding missions envisioned for beam weapons.

One of driving factor for designing a good fire control system for DEW is the mission requirements assigned to such weapon system. The different missions proposed for target engagement of beam weapons impose distinct requirements for each of these missions and as a result on beam and its fire control system as well as RSTA couple with DEW platform. In general, the mission can be roughly divided into two categories that are known by early definition of strategic defense intuitive (SDI) threats. The two groups are:

1. **Endoatmospheric** threats and target engagement within the atmosphere
2. **Exoatmospheric** threats and target engagement within the outer space, where there is no air to deal with such as thermal blooming and beam divergence effects or encountering particles in the air

Although some proposed systems fall into a hazy intermediate category because they require sending a laser beam up from the ground into the space or look down and shoot down from space in boost phase of incoming threat such as continental ballistic missile.

Directed energy weapons intended for use in the atmosphere typically have tactical missions on or above the battlefield. That means that they must be able to function over ranges measured in miles or kilometers. For all practical purposes, a beam traveling at or near the speed of light can reach such a target instantaneously. However, the atmosphere can bend, distort, or break up the beam, and sophisticated compensation techniques are needed to concentrate the beam energy onto the right point of the target. Many missions, such as defense of a battleship against an onslaught of cruise missiles, would probably require extremely fast response to engage and zap many targets before they could reach the designated target of their mission (i.e., ships). Speed might not be as critical in some other missions, such as destroying comparatively slow-moving helicopters, but in most cases, the system would have to pinpoint enemy targets in a field that included friendly forces. Typical targets would require concentration of enough power at a critical point to physically disable them, but in some cases, much lower powers would be sufficient to blind a sensor and thereby disable the target without any damage to friendly force [i.e., *bidirectional reflectance distribution function* (BRDF)] effects (see Chap. 2 on Laser Safety) [2].

Ground- and air-based anti-satellite (ASAT) weapons present rather different design constraints. Their range would be hundreds or perhaps thousands of kilometers or miles, depending on the orbit of the target satellite. Moderate powers reaching the satellite should be sufficient if the goal was to blind sensors or disable vulnerable sensing systems or electronics, probably the most effective way to kill current satellites. Seen from the ground, satellites would be slow-moving targets. Some compensation for atmospheric effects (i.e., thermal blooming) might be needed, but the task would be easier than if high powers had to be delivered to a small point on the target.

Considering that if we divide the target acquisition and engagement into three stages—(1) initial course or boost phase (endoatmospheric scenario), (2) - mid-course phase for incoming ballistic missile (exoatmospheric scenario), and

(3) final course (endoatmospheric scenario)—then defending the country against a nuclear attack requires a different system and would have a very different set of requirements. In an all-out attack, hundreds of targets would simultaneously appear thousands of kilometers or miles away, and the weapon system would have to go after them as rapidly as possible at that distance. Except for X-ray lasers, any weapon system would have to destroy a series of many targets in succession. The degree of devastation caused by a nuclear bomb would make it important to kill as many targets as possible (ideally all of them) and to know which ones were not disabled so other weapons could shoot at them. If the lasers were in space, there would be no need to worry about atmospheric effects at all. If the lasers were on the ground, transmitting their beams to "battle mirrors" in space, compensation for atmospheric effects could be performed both by the optics on the laser and by the large mirrors in space, which together would form the beam control system.

Additional differences come from the nature of the weapon itself. It takes different techniques to direct beams of visible light, X-rays, charged particles, uncharged particles, and microwaves. Fire control techniques for such weapons would be more closely related to each other, although there would be some notable differences. Our focus in this chapter is based on the high-energy laser weapon system. The different ways in which other types of beams would be controlled will be covered in Chap. 7 and 8 of this book, which describes the principal directed energy alternative to lasers in more details.

6.2 Directed Energy Weapons Engagements

Target engagements with directed energy weapons (DEWs) are very likely to be considerably different from conventional engagements with kinetic weapons. They will demand a much more detailed knowledge of the specific target engaged, and it will be harder to perform damage assessment during and after the engagement.

First, the target knowledge required for a DEW engagement will most likely be significantly different and much more detailed than that required for a kinetic engagement. The variation in effectiveness of DEWs against multiple target sets is likely to be quite large. It is not likely that weapon developers can produce devices that will be as universally effective as typical kinetic weapons. It will be important for DEWs to be employed against specific components in most targets, and those components need to be identified prior to the engagements.

Second, our ability to perform damage assessment after a DEW engagement is likely to be significantly different than currently used methods. Physical damage may not be easily observable. Anomalous behavior may be the only clue to a directed energy weapon's effectiveness on a specific target. A simple example would be that of an air-to-air engagement. With a kinetic weapon, an attack on an aircraft will be recognized as successful if the aircraft bursts into flames or it dives uncontrollably toward the Earth. Both observations will indicate a successful kill. With a DEW system, the attack may never produce a physical change in the target

that is observable, only the uncontrolled behavior of diving toward the Earth would indicate a successful kill. Attacks on ballistic missiles in the boost phase may well produce easily observed explosions as well as anomalous behavior as a result of irradiation.

The details of how these two factors are handled will vary with the type of DEW, whether it is a laser device, a high-power microwave device, or a particle beam device.

Consider first the laser weapon. Any laser system is likely to be fairly expensive, though it may have a low cost per target engagement. Therefore it will be desirable for the laser system to be able to engage targets outside of their kinetic kill range. It should be fairly easy for a laser to do this, and it is one of the advantages of this type of DEW that it will have a long range of effectiveness by conventional standards.

Consider the following table for the potential spot sizes generated by laser weapons operating at 1.0 and 10.0 μm with diffraction-limited beams.

Range	Spot size radius at 1.0 m	Spot size radius at 10.0 m
100 m	0.1 mm	1.0 mm
1 km	1.0 mm	10.0 mm
10 km	1.0 cm	10.0 cm
100 km	10.0 cm	100.0 cm
1000 km	1.0 m	10.0 m
10000 km	10.0 m	100.0 m

Of course optics will never be perfect, and the atmosphere (thermal blooming effects, etc. in case of laser beam weapons) will perturb the beam significantly, but it would appear that engagements out to 1–10 km are possible.

At these ranges, a typical radar return is probably not adequate to locate a target accurately enough to commit to firing. Therefore a tracking laser of some sort will be required to accurately locate the target and aim the high-power laser. The tracking laser will also be required to image the target as the spot sizes at these ranges are small enough that a specific aimpoint on a target can be selected to maximize the beams effect. For this to be accomplished, a reasonably detailed knowledge of the target geometry must be available. Boresighting the tracking laser and the high-power laser will require great precision. For actual laser weapon systems, the acquisition, pointing and tracking challenges are at least as great as the beam generation challenges.

To get a feel for the issues involved, it is probably useful to discuss the actual successful target engagements accomplished by the Airborne Laser Laboratory (ALL) in 1983 [1]. The ALL used a 10.6-μm laser to engage two targets, an AIM-9B Sidewinder missile and a BQM-34A Aerobe drone in separate test series. The AIM-9B test series addressed the aircraft self-defense scenario against an air-to-air missile. The BQM-34A test series addressed the fleet defense scenario against a cruise missile.

The first two AIM-9B engagements started at approximately 3-km range and had beam on times of 4.8 and 3.8 s. The beam spot sizes were estimated to be less than

10 cm in radius. In both engagements, the tracking laser acquired the missile and aimed the high-power laser so as to hit the missile. The returned glint off the body of the missile allowed the beam to be walked forward to the nose section containing the tracker head of the missile guidance system. The beam then dwelled on the tracker head resulting in burn through of the nose dome. The actual burn through took less than a second. When the tracker head failed, the missile veered off course and crashed into the ground. The entire engagement was observed on the tracking screen in the ALL fire control center. The engagement was computer controlled as the response times and tracking maneuvers exceeded any human capability. Kill confirmation occurred when the missile veered off course. However, the AIM-9B was instrumented in such a fashion that burn through of the nose dome and failure of its tracking mechanism were transmitted to an instrumentation aircraft following the ALL to confirm the effects of the laser irradiation. Still there were no visible indications of missile kill comparable to a kinetic kill and subsequent explosion of the missile.

So for the third engagement, a small spotting charge was rigged to a break-wire in the nose of the missile. It would explode if the wire burned through due to laser irradiation. This time the scanning, movement of the beam to the nose of the missile, and burn through of the dome took 2.4 s, and the missile exploded with a puff of white smoke. This provided an additional confirmation of missile kill suitable for publication in the press but of no consequence for system validation. The fourth and fifth engagements required 3.6 s and 3.1 s, respectively, to observe the AIM-9B's fatal veering off course and crash.

The engagements were successful because the computer on the ALL had enough information on the return from the AIM-9B's to know the most likely strongest glint return from the Sidewinders body and the offset from this point to the nose dome of the missile. The engagements could also be terminated with a reasonable beam on time because the veering off course was a clear indication that the missile's guidance system had failed. All tracking was accomplished by computer, and no visible laser beam was ever observed. The engagements were significantly different than conventional engagements, but just as successful in terms of destroying the target as kinetic kills would have been.

For the engagements of the BQM-34A, an aimpoint just forward of the wing root was chosen to install a stainless steel tank to simulate a fuel tank on a cruise missile. The glint from the nose of the drone was used as the reference point for tracking, and the aimpoint was offset from that reference. On the first ALL engagement, the laser hit the tank and dwelled long enough to detonate some fuel within the tank, but the explosion was not strong enough to destroy the drone. The drone ran out of fuel and crashed into the ocean. In the second ALL engagement, the beam control allowed the beam to shift off the fuel tank and drift down to the wing root. This drone was recovered from the ocean and had obvious damage to the wing root. But the damage was not enough to cause the drone to veer out of control and destroy itself. (One could draw an analogy to the vehicle being hit by small firearms and not being destroyed.) On the third ALL engagement, the aimpoint was shifted from the fuel tank to the area of the fuselage containing the flight control system, slightly

behind the wing. At a range of 1.8 km, the ALL placed the beam on the surface above the flight control box. Almost immediately the flight control signals being telemetered to the tracking aircraft were disrupted. The drone rocked rapidly back and forth, and then it took a hard 90° roll to the right and dove into the water. Once again kill confirmation was obtained by anomalous behavior of the target.

Based on these two sets of kills by an engineered high-power laser system, future systems will require:

1. Detailed target knowledge, if the target is larger than a few tens of centimeters
2. Computer-controlled tracking utilizing multiple lasers and recognition of laser return signals from a spatially diverse target
3. Kill confirmation by anomalous target behavior

These requirements appear achievable, but they are somewhat different than kinetic system requirements.

Not all laser engagements are anticipated to occur against aircraft or missile targets. Engagement of ballistic targets is not likely to give indications that would be useful for kill confirmation. If a kill cannot be easily confirmed, then it becomes difficult to determine when to terminate dwell time on the target. Engagement of fixed targets on the ground suffers from a similar lack of kill confirmation mechanism.

High-power microwave (HPM) weapons or electromagnetic pulse weapons suffer from the same difficulties to some extent. Aiming an HPM weapon is much easier than a laser because the radiation is far less collimated and in many cases may almost be isotropic. This however presents a major limitation to employment, as the device must get much closer to its target before emitting. It also presents problems of "suicide" or "fratricide" to the weapons delivery system or systems.

When it comes to target knowledge for an HPM system, there are classically two types of effects or couplings into the target. In front door coupling, the attacking system uses an antenna that is tuned for a specific frequency or range of frequencies to overpower the receiver of the target. This is usually a solvable problem and the power on target can be estimated. However, front door coupling is only useful against systems having a receiving antenna that is accessible and necessary for its continued operation. This applies primarily to communication systems.

In back door coupling, the attacking system attempts to couple into the target through some method that is out of band and/or does not come in through an antenna. This is a much more difficult problem and pretty much not capable of being solved by analysis. Simply put, an attacker must have available samples of systems that will be attacked and be able to test for the effects desired prior to carrying out an attack. There are many ways that an electromagnetic signal can couple into a circuit inside a porous body, but accurate analysis is extremely difficult to perform from first principles. Shielding against inadvertent entry of electromagnetic signals can be accomplished, and systems can be hardened to the effects of HPM, but maintaining such shielding is difficult and probably not very

cost-effective. To validate that an attacker has a high probability of success against a particular system, many tests are required.

Finally, electromagnetic kills or upsets are very difficult to verify. If a communication or radar system is attacked, its failure to operate can be observed electromagnetically. Beyond that, confirmation of electromagnetic damage effects is very difficult. It is very likely that the success of such devices will have to be estimated based on lots of tests and certain risks accepted. Future research may develop active interrogation schemes that can assess the effects of an HPM attack, but there are few such interrogation schemes currently available.

Particle beams are the least developed of DEWs, and no known system test to determine operating parameters for a successful kill has been determined. It is highly likely that a similar level of target information will be required for the employment of particle beams as is required for high-power lasers. Since particle beams do penetrate their targets, the interior designs may be necessary to identify aiming points and kill mechanisms. The effects of particle beams are better defined on a geometric configuration of materials than electromagnetic waves, and many effects can be calculated from first principles if the materials and geometries are known.

Both electron beams and proton beams will produce a target return of electromagnetic radiation in the form of X-rays and gamma rays. Protons will also produce a return of neutrons. The intensity of these returns will vary with the target that is intercepted. Relative or spectral intensities may also provide some information on target materials. Whether these will be adequate for steering an invisible beam in space remains to be seen. Certainly lasers can be used to assist in the pointing and tracking problem.

A successful kill of a ballistic warhead by a particle beam will be difficult to identify unless enough energy is deposited to cause the target to self-explode. This may require a prohibitive amount of energy in the beam. If the goal is an electronic kill, any method developed for determining electronic kills for HPM weapons will be useful. Kills of aircraft or maneuvering systems in the atmosphere will be identified by the same anomalous behavior as has been identified for high-power lasers.

6.2.1 Acquisition, Tracking, Pointing, and Fire Control

Directing the laser energy from the optics to the target requires a highly accurate acquisition, tracking, pointing, and fire control system. A laser weapon system, either space based or ground based, needs to locate the missile (acquisition), track its motion (tracking), determine the laser aimpoint and maintain the laser energy on the target (pointing), and finally swing to a new target (fire control). The accuracy for each component is stringent because of the great distances between the weapon and the targets.

The United States put considerable time and resources into both space and ground programs in acquisition, tracking, and pointing technologies. Space experiments are critical to any high-energy laser weapon system because they demonstrate the high-risk technologies and do so in the actual operational environment. However, the space programs in the 1980s suffered from high costs and the space shuttle challenger accident while many programs were terminated or had their scope reduced due to insufficient funding, two highly successful space experiments were completed in 1990. The Relay Mirror Experiment demonstrated the ability to engage in high-accuracy pointing, laser beam stability, and long-duration beam relays. This is a critical technology for any weapon architecture that requires relay mirrors in space. Another successful test was the Low-Power Atmospheric Compensation Experiment that was conducted by the MIT Lincoln Laboratory, which demonstrated the feasibility of technologies that are designed to compensate for the atmospheric turbulence that distorts laser beams.

A number of the space experiments were canceled or redesigned as ground experiments. Ground experiments can be successfully conducted as long as the tests are not limited or degraded by the Earth's gravity. Two ground experiments demonstrated the key technologies that are essential for the space weapon platform to maintain the laser beam on the target despite the large vibrations induced by the mechanical pumps of a high-energy chemical laser [3]. The rapid retargeting/precision pointing simulator was designed to replicate the dynamic environment of large space structures. Using this technology, which is especially critical for a space-based laser, scientists tested methods to stabilize the laser beam, maintain its accuracy, and rapidly retarget. Within the constraints of a ground environment, the techniques developed should be applicable to space systems.

Another successful experiment was the Space Active Vibration Isolation project, which established a pointing stability of less than 100 nanoradians. This equates to 4 in. from a distance of 1000 km. The Space Integrated Controls Experiment followed that program and further improved the pointing stability 3. To understand the technology necessary to control large structures, such as space mirrors, the Structure and Pointing Integrated Control

Experiment Stratospheric Particle Injection for Climate Engineering (SPICE) was developed to demonstrate the value of active, adaptive control of large optical structures. These tests, experiments, and demonstrations represent the current state-of-the-art laser technology, which leads to the question of how to fit these technologies into architecture and how much further to push the technology.

6.3 Wavelength Effects

The wavelength of a laser sets some fundamental constraints on the optics that can be used with the laser. By far the most important is the Fraunhofer diffraction limit, which determines how small a spot the beam can form. In this case, the spot size is measured as an angle (as viewed from the laser) that is proportional to the ratio of

wavelength to the diameter of the focusing optics. The theoretical formula for the ideal case is [4]

$$\text{Spot Size} = 1.22 \times \frac{\text{wavelength}}{\text{optics diameter}} \tag{6.1}$$

This formula actually defines the first point at which the intensity falls to zero. There are a series of bright rings surrounding the control spot, falling off in intensity as the distance from the central spot increases. The formula is valid for a circular output mirror and a perfectly uniform laser beam, which does not exist in practice but which does give a rough approximation for real lasers. It actually gives the sine of the angular spot size for small spots such as would be produced by a laser beam, which is virtually identical to the angle in radians. The formula is a fundamental one and comes from Donald H. Menzel [4].

This formula gives spot size in radians, an angular measure equal to the diameter of the spot divided by the distance to it. (One radian equals 57.3°.) Laser physicists usually talk in terms of spot size (sometimes sloppily called beam divergence, although in this sense it isn't exactly that) in radians. However, the formula can be altered to give spot size in meters:

$$\text{Spot Diameter (meter)} = \frac{1.22 \times \text{wavelength} \times \text{target distance}}{\text{optics diameter}} \tag{6.2}$$

Note that in making calculations with either formula, it is essential that everything be measured in the same units. Thus, if wavelength starts in micrometers, target distance in kilometers, and optics diameter in meters, they would all have to be converted to meters before making the calculation. Spot diameter is important because it indicates onto how small an area on the target the laser's output can be concentrated. Dividing the area of the focal spot on the target into the laser power or energy gives the power or energy density at the target. Measurements of laser power or energy density on the target are useful in making rough approximations of the threshold for causing damage, but the actual mechanisms involved are quite complex and far beyond the scope of this book.

Another important factor is the beam "wander" or "jitter." That is, how precisely can the laser beam be kept on one spot on the target while it is depositing its lethal dose of energy? If the beam wanders all over the place, it will not stay at any one point long enough to do any damage. The usual assumption is that beam wander will have to be somewhat smaller than the spot size.

One suggestion to correct such wandering or jitter is coming from P. Sprangle, A. Ting, J. Peñano, R. Fischer of Plasma Physics Division of Naval Research Laboratory (NRL), and B. Hafizi of Icarus Research Inc in their paper *High-Power Fiber Lasers for Directed-Energy Application*. Their recommendation for compensating for wandering or jittering of laser beam is an approach known as "beam wander and tip–tilt compensation".

Introducing tip–tilt correction into the individual steering mirrors can reduce the overall laser spot size on target. Tip–tilt correction redirects the centroid' of the individual laser beams to cancel the effects of wander due to turbulence. This is accomplished by monitoring the intensity on target and redirecting the steering mirrors to minimize the spot size. Laser beam wander is a function of the scale size of the turbulence fluctuations. Turbulent eddies that are large compared to the laser beam diameter cause the laser beam centroid to be deflected and to wander in time due to transverse air flow. Eddies that are much smaller than the beam diameter cause spreading about the beam centroid and cannot be reduced by the use of tip–tilt compensation. The observed long-time averaged laser spot size is a combination of beam wander and spreading about the centroid. In weak turbulence, the beam centroid wander represents a significant contribution to the laser beam radius. As the turbulence level increases, or for long propagation ranges, the beam wander contribution to the laser spot size becomes less important. In very strong turbulence, the laser beam breaks up into multiple beams making tip–tilt compensation ineffective. If the individual laser beams are separated by less than r_0 at the source, the wander of the centroids on the target will be correlated. In this case, it would be possible for beams to share a common tip–tilt correcting aperture, thus reducing the size and complexity of the system.

With this information in mind, you can calculate some very general requirements on a laser weapon system if you know the lethal power or energy density and the maximum range. For example, suppose that in the case of missile defense, it seems desirable to concentrate a laser power of about 5 million watts onto a spot about 1 m (about a yard) in diameter (a little over 6 million watts per square meter). If the target is 5000 km (3000 miles) away, in angular terms, the focal spot is 0.2 millionths of a radian (or 0.2 μrad in standard scientific terminology). That figure can be inserted into the equation that relates laser wavelength and optics diameter to spot size. Suppose, for example, the weapon system uses a space-based hydrogen fluoride chemical laser with a nominal wavelength of 2.8 μm or 0.0000028 m. Simple division shows that the output mirror must be 17 m (56 ft) in diameter. If that doesn't sound impressive enough, you should realize that the largest telescope in the United States, the 200-in. giant at Mount Palomar Observatory, has a main mirror of only 5 m in diameter. The largest telescope mirror in the world is a 6-m one in the Soviet Union, which unofficial sources report hasn't been working very well. The largest mirror yet designed and built for use in space is the 2.4-m (8-ft) mirror for NASA's space telescope.

The use of a laser with higher-power or shorter wavelength would allow use of a smaller mirror. If a 1.3-μm chemical oxygen iodine laser could be substituted for the hydrogen fluoride laser in the previous example, a mirror 8 m (26 ft) in diameter could produce a 1-m spot 5000 km (3000 miles) away. Increasing the laser's output power does not produce such dramatic reductions in required mirror size because damage depends on power density multiplied by illumination time, which increases faster with decreasing spot size than with increasing power. Thus, a 10 million-watt beam could be spread over twice as much *area* as a 5 million-watt beam, but as the area doubled, the spot diameter would increase only by the square root of 2, a factor

of 1.4. Thus, a 10 million-watt hydrogen fluoride laser beam could be focused onto a spot 1.4 m in diameter at a distance of 5000 km (3000 miles) with a mirror only 12 m (40 ft) in diameter, yielding the same power density as would be obtained when focusing a 5 million-watt laser over the same distance with a 17-m (56-ft) mirror onto a 1-m spot.

From a practical standpoint, shrinking the mirror diameter from 17 m (56 ft) to 12 m (40 ft) would be important. On the ground or in space, the weight of a massive mirror would present a problem. If weight was simply proportional to area (which in turn depends on the square of the diameter), reducing the diameter from 17 to 12 m would cut the weight in half. In practice, an even greater weight reduction would be possible, because the smaller-diameter mirror could be thinner and still have enough mechanical strength to maintain its shape [2].

These simple calculations demonstrate why The Pentagon is so interested in developing short-wavelength lasers. The prime allure of the ultraviolet is a wavelength about one-tenth that of the hydrogen fluoride laser, making it possible to use a mirror much smaller than needed for an iodine laser, although other considerations described below might weigh against picking the smallest possible mirror diameter [2].

It may be possible to put a laser weapon for use against missiles or satellites on the ground, but if the targets were in space, the focusing mirror would have to be there, too. Putting a large mirror into orbit is not going to be an easy job. Current proposals for space-based lasers envision either a 5 million-watt chemical laser with a 4-m (13-ft) mirror or a more potent weapon using a 10 million-watt laser and a 10-m (33-ft) mirror. Mirror sizes would have to be the same to deliver the same power to the same size spot, with the same laser wavelength, even if the laser was on the ground. Military contractors seem to think the task is achievable. The Corning Museum of Glass, PerkinElmer Corp., Itek Corp., and Eastman Kodak have proposed a plan for a 4-m (13-ft) glass mirror. The United Technologies Research Center has offered to build a 10-m (33-ft) lightweight mirror, using a graphite fiber-reinforced glass matrix for the body of the mirror and vaporized silicon for the reflective coating (Fig. 6.1).

The mirrors best able to stand up to high laser powers are made of solid metal typically honeycombed with holes through which coolant flows. The Department of Defense has spent millions of dollars developing ways to produce mirrors that absorb less than 1 % of the incident laser light and which I can efficiently conduct away what heat they do absorb. Much of the effort has gone to development of machines that use diamond-edged tools to cut mirror surfaces directly into metal blocks, vastly simplifying the traditional time-consuming process of making optics. Diamond turning, as the technique is called, also makes it possible to produce mirrors with surface shapes impossible to obtain by conventional grinding and polishing methods [2].

For more details on this subject, please refer to the book by Jeff Hecht “Beam Weapons, The Next Arms Race.”

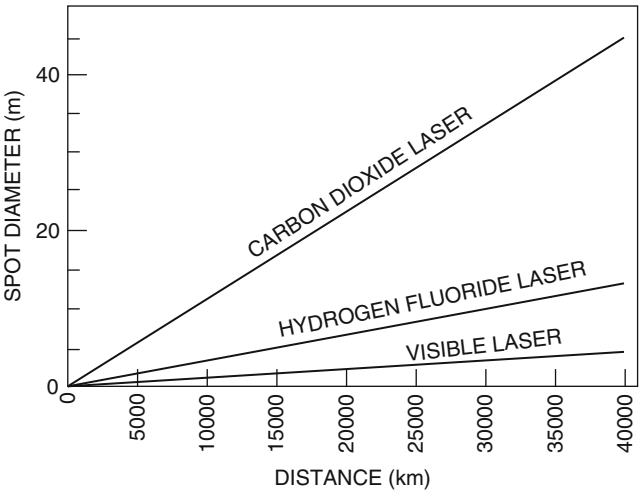


Fig. 6.1 Laser spot size that would be at various ranges and wavelengths by a 30-m output mirror. The graphs show spot diameter for the 10.6-μm carbon dioxide laser, the 2.8-μm hydrogen fluoride laser, and a hypothetical 0.5-μm visible laser, assuming that the figure of the mirror is accurate to within 1/50th of the laser wavelength. (Drawing by Arthur Giordani based on calculations by Wayne S. Jones, Lockheed Missiles and Space Co.)

6.4 The Atmospheric Propagation Problem

Air looks much more transparent than it really is. In the past decade, researchers are working on a communication systems that rely on laser beams going through the air. New technologies in direct line of sight wireless communication systems are being developed at the University of Maryland. These are hybrid systems that use both free space optical (FSO) communication links and radio frequency (RF) links [5]. In order to optimize these systems, an understanding of the channel medium (the atmosphere) is requisite. The concern of the current research is the characterization of the effects of atmospheric turbulence on the propagation of electromagnetic waves at optical wavelengths.

One effect of atmospheric turbulence is that it scatters light. The current method of research into the characteristics of atmospheric turbulence is the analysis of this scattered light. It is this resultant scattered light that brings up two essential questions. What does the turbulence do to the light, i.e., how does it scatter the light? What does the scattered light tell us about the atmospheric turbulence?

The answer to the first question is important for engineering applications like an FSO communication systems. Specifically our current research will look to optimize the receiver or aperture size for an FSO communication system. The answer to the second is important to the development of formal mathematical theories on the behavior or physics of the atmosphere.

The effects of atmospheric turbulence are easily seen by the propagation of the laser through the atmosphere. As can be seen in Fig. 6.2, the turbulence induces

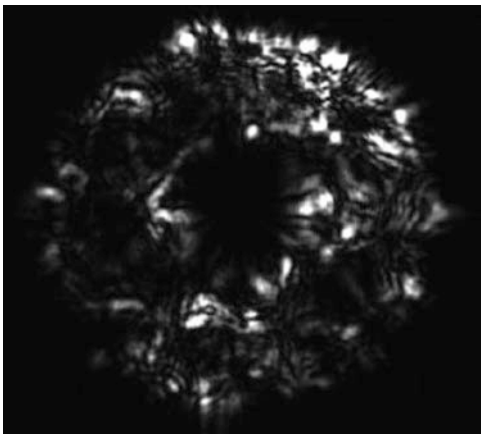


Fig. 6.2 Screen shot of image [5]

intensity random fluctuations in the beam known as scintillation. This is an effect of the many random changes in the index of refraction along the path of beam propagation due to turbulence.

Their current research will aid the development of these physical theories where they assumed that the atmosphere can be described as a moving fluid of air. As such it can be modeled by the Navier–Stokes equations (shown below) that describe the behavior of moving fluids [5]:

$$\begin{gathered}\frac{\partial \mathrm{v}}{\partial t}+(v\cdot\nabla)v=-\frac{1}{\rho}\nabla p+\upsilon\nabla^2 v\quad \nabla\cdot v=0\\ \text{Navier}-\text{Stokes Equations (NSE)}\\ \text{(Assuming incompressibility)}\end{gathered}\tag{6.3}$$

where ρ is density, $\upsilon=\frac{\eta}{\rho}$ is called the kinematic viscosity, and η is called the dynamic viscosity.

These equations show that atmosphere can move as either as turbulent flow or laminar (nonturbulent) flow. Turbulence occurs when a fluid flow exceeds a critical Reynolds's number $R\equiv lu/\upsilon$ with l being a characteristic linear dimension (traveled length of fluid or hydraulic diameter when dealing with river systems) (m) while u is the mean fluid velocity (SI units: m/s) which causes the nonlinear term $((v\cdot\nabla)v)$ of the NSE to dominate which is characterized by the flow's chaotic behavior (as shown below):

$$\frac{(v\cdot\nabla)v}{\frac{\eta}{\rho}\nabla^2 v}\sim\frac{u^2/l}{\eta u/\rho l^2}=\frac{ul}{\upsilon}=R\tag{6.4}$$

The nonlinear nature of the equations makes them difficult to with since there are no known solutions or families of solutions to the Navier–Stokes equations. This forces one to look at other mathematical observations to take something meaningful

from these equations. In low turbulence (when the nonlinear term, $(v \cdot \nabla)v$, is much less than $\upsilon\nabla^2 v$), significant symmetries arise. These symmetries are space and time translation, Galilean transformation, parity, rotation, and scaling. In extreme turbulence, Kolmogorov predicted and proved that the symmetries of the Navier–Stokes equations that exist in nonturbulent systems reappear. The extreme complications of the Navier–Stokes equations, arising from the nonlinearity, force us to use statistical methods to characterize the behavior of a turbulent fluid as such the atmosphere. Fortunately, the symmetries give us insight into the statistics.

High-energy lasers (HELs) have a number of directed energy (DE) applications [6] requiring high-intensity beams to be propagated at long distances under a wide range of atmospheric conditions. The optimum wavelength for efficient HEL propagation depends on the atmospheric conditions and a number of interrelated physical processes which include thermal blooming due to aerosol and molecular absorption [7], turbulence [8], aerosol and molecular scattering [9], thermal scattering due to heated aerosols, and aerosol heating and vaporization [3, 10–15]. The relative importance of these processes depends on the parameters of the atmospheric environment which can vary significantly depending on location and time. The main objective that was discussed by Phillip Sprangle et al. [6] is about the optimum laser wavelength and power for efficient propagation in maritime, desert, rural and urban atmospheric environments. The theoretical/numerical model used in this study includes the effects of aerosol and molecular scattering, aerosol heating and vaporization, thermal blooming due to aerosol and molecular absorption, atmospheric turbulence, and beam quality. These processes are modeled in a fully three-dimensional and time-dependent manner. It is found that aerosols, which consist of water, sea salt, organic matter, dust, soot, biomass smoke, urban pollutants, etc., are particularly important because they result in laser scattering, absorption, and enhanced thermal blooming. In the water vapor transmission windows, the total absorption coefficient driving thermal blooming can be caused mainly by aerosols and not water vapor. In certain maritime environments, the deleterious effects of aerosols can be reduced by vaporization. Aerosols which cannot be vaporized, such as those consisting of dust, soot, etc., can significantly increase thermal blooming. We show that moderate values of the laser beam quality parameter have little effect on the propagation efficiency. The laser power, averaged over dwell time, delivered to a distant target as a function of transmitted power is obtained for a number of wavelengths and atmospheric environments. The optimum wavelength and power are found for each atmospheric environment.

Atmospheric environments contain various types and concentrations of aerosols which can, for HEL beams, enhance thermal blooming and significantly affect the propagation efficiency. In general, aerosols consist of hygroscopic and non-hygroscopic particles of various sizes and chemical compositions. Hygroscopic aerosols are water soluble and vary in size depending on the relative humidity [15]. Oceanic aerosols consist of sea salt, water, and organic material. Non-hygroscopic aerosols are composed of dust, soot, biomass smoke, and other

carbon-based compounds. These aerosols typically have much larger absorption coefficients than water-based aerosols. While they are normally present in continental, rural, and urban environments, dust aerosols can also be present in maritime environments hundreds of miles from the shore [16].

Aerosols can absorb laser energy, and, in the case of hygroscopic aerosols, the absorbed energy goes into both heating and vaporizing the aerosol. Heated aerosols conductively heat the surrounding air, resulting in an increase in thermal blooming of the HEL beam. However, since aerosol scattering and absorption coefficients are strongly dependent on the aerosol radius, vaporizing the aerosol can improve the propagation efficiency. Non-hygroscopic aerosols (dust, soot, etc.), however, have large scattering and absorption coefficients and will not vaporize at the intensity levels anticipated in DE applications. These aerosols continually heat the surrounding air leading to significant thermal blooming. Water vapor absorption bands and those of carbon dioxide determine the atmospheric transmission windows in the infrared. Under a range of atmospheric conditions and laser wavelengths, aerosol absorption can exceed water vapor absorption and thus can be the dominant process for thermal blooming. For example, in a maritime environment at an operating wavelength of $\lambda = 1.045\,\mu\text{m}$, the water vapor absorption coefficient is $\sim 3 \times 10^{-5}\text{km}^{-1}$, while the aerosol absorption coefficient is often greater than 10^{-3}km^{-1}. In other water vapor transmission windows, i.e., 1.625 and 2.141 μm, the water vapor and aerosol absorption coefficients can be comparable. In addition to enhancing thermal blooming, aerosols can also significantly contribute to the total laser scattering coefficient [6].

Longer wavelength, such as microwaves, can penetrate through dust and precipitation because the particles in the air are much smaller than the wavelength. In some cases, longer-infrared wavelengths are transmitted much better than visible light because of the difference in their wavelengths.

Under typical atmospheric conditions, the absorption of 10.6 μ radiation by various molecular species results in an absorption coefficient of approximately $1.5 \times 10^{-6}\text{cm}^{-1}$ and leads to various nonlinear propagation problems [17, 18]. For some lasers (e.g., the DF laser operating at 3.8 μ), the molecular absorption in the air is small (as low as 10^{-8}cm^{-1}), and the heating due to aerosols can dominate the heating of the air path. For a typical atmosphere, this aerosol absorption corresponds to an attenuation coefficient of approximately 10^{-7}cm^{-1} but could be higher if the aerosol concentration was unusually high [19]. In general, the nonlinear heating effects are more important than the linear power losses.

The natural heating effects of high-energy laser beams could help to overcome some weather-related transmission problems. A high-energy carbon dioxide laser, for example, can bore its way through fog by heating the tiny water droplets that obscure vision enough to make them evaporate. Much the same could be done with clouds. Heavy rain would be harder, particularly in the realistic situation where the beam is being scanned quickly through rapidly falling rain. If the beam scans across the drops too quickly, it may only partly evaporate them before moving on to illuminate other drops.

Foul weather, dust, and smoke can do more than just blocking the beam from the high-energy lasers. They can make it impossible to find targets visually or with infrared optical system, making it necessary to rely on microwave radar that can penetrate the obscuration. However, even if the high-power beam could burn its way through, the limited resolution of microwave radars may not be able to pinpoint the target accurately enough for the laser beam to hit a vulnerable spot.

6.4.1 Laser Light Scattering and Intensity

In general, a laser beam is attenuated as it propagates through the atmosphere. In addition, the laser beam is often broadened, defocused, and may even be deflected (i.e., scattered) from its initial propagation direction. These atmospheric effects have far reaching consequences for the use of laser in optical communication, weaponry, target designation, ranging, remote sensing, and other applications that require transmission of laser beams through the atmosphere. In a tactical battlefield, a new spectral light transport model for sand and other obstacles in the air should be studied and be mathematically analyzed and experimentally put it into test. The model should employ a novel approach to simulated light interaction with particulate materials which yields both the spectral and spatial bidirectional reflectance distribution function (BRDF) responses of sand or other obstacles in the environment of battlefield which includes natural and man-made media such as rain or fog and gun smoke. Furthermore, the parameters specifying the model should be based on the physical and mineralogical properties of sand and other obstacles of the concern. The model should be evaluated quantitatively, through comparisons with measured data. Good spectral reconstructions if any should be achieved for the reflectances of several real sand samples. Its potential applications include, but are not limited to, applied optics, remote-sensing and image synthesis, as well as weaponry.

Light scattering has provided an important method for characterizing macromolecules for at least three decades. However, the replacement of conventional light sources by lasers in recent years has qualitatively changed the field and has sparked renewed interest. Through the intense, coherent laser light and efficient spectrum analyzers and autocorrelators, experiments in the frequency and time domains can now be used to study molecular motion and other dynamical process. Classical light scattering studies are concerned with the measurement of the intensity of scattered light as a function of the scattering angle. In addition to this kind of study, laser light sources now permit spectral information (i.e., BRDF) to be obtained from the scattered light. The latter type of experiment is often called quasi-elastic light scattering (QLS), and the various forms of the experiment are known as light beating spectroscopy (LBS), intensity fluctuation spectroscopy (IFS), and photon correlation spectroscopy (PCS). Related experiments in laser Doppler velocimetry (LDV) now permit very low rates of uniform motion to be measured.

A special case of LDV is electrophoretic light scattering (ELS) where motilities are determined [20].

The attenuation and amount of beam alteration depend on the wavelength, output power, makeup of the atmosphere, and the day-to-day atmospheric conditions. When the output power is low, the effects are linear in behavior. That is, doubling the initial beam intensity results in a doubled intensity at every point along the beam's path. Absorption, scattering, and atmospheric turbulence are examples of linear effects. On the other hand, when the power is sufficiently high, new effects are observed that are characterized by nonlinear relationships. Some important nonlinear effects are thermal blooming, kinetic cooling, beam trapping, two-photon absorption, bleaching, and atmospheric breakdown, which, incidentally, fix an upper limit on the intensity that can be transmitted. In both cases, the effects can be significant and may severely limit the usefulness of the beam. Many studies have been undertaken to define the different linear and nonlinear phenomena that can occur [21].

One parameter that is useful in determining the effectiveness of a Gaussian laser beam is the beam irradiance at the target. For a beam with output power P_0 and cross-sectional area A at the target, the peak irradiance I_p at the target is [21]

$$I_p = \frac{P_0 \tau}{A} \tag{6.5}$$

where τ is the atmospheric transmittance. A major system design goal is to maximize I_p by minimizing the beam cross-sectional area A and maximizing the product $P_0\tau$.

The propagation of a laser beam in a vacuum is governed by the diffraction theory, which tells us that no matter how parallel the beam may be initially, it will diverge and spaced as the beam propagates away from its source. Most laser beams have a Gaussian intensity profile in the transverse front. The beam radius w is defined as that transverse distance from the center of the beam (i.e., the beam axis) to the point where the intensity has fallen to $1/e^2$ (0.13533) of its on-axis value (see Fig. 6.3).

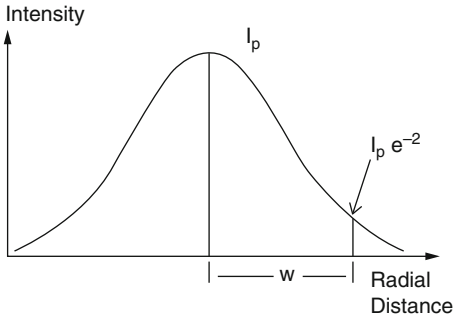


Fig. 6.3 Intensity profile of Gaussian beam

The two parameters of most interest in describing the propagation characteristic of a Gaussian beam are the beam radius $w(z)$ also known as spot size at any distance z from the beam waist and the radius of curvature of the phase front $R(z)$. These two parameters are given by

$$w(z) = w_0\sqrt{1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2} \tag{6.6}$$

and

$$R(z) = z\left[1 + \left(\frac{\pi w_0^2}{\lambda z}\right)^2\right] \tag{6.7}$$

where w_0 is the radius at the beam waist (that part of the beam where the beam has its smallest diameter) and λ is the wavelength [21].

Like other beams, Gaussian beams diverge as they propagate through space. However, the intensity distribution remains Gaussian in every beam cross section. Only the width of the Gaussian profile increases as the beam propagates. At the beam waist ($z = 0$ and $R = \infty$), the phase front is a plane. For most practical lasers, the beam waist is located a short distance from the output mirror on the outside of the resonator (see Fig. 6.4). As the Gaussian beam propagates away from the location of its waist, the beam's radius (or spot size) at first remains nearly constant but then begins to diverge linearly with distance at large distances from the waist. The smaller the spot size at the waist, the faster the beam diverges and the smaller the distance over which it stays collimated with a near constant diameter and a near planar wave front. The angle θ in Fig. 6.4 is the beam divergence angle. It is given by

$$\theta = \frac{w(z)}{z} = \frac{\lambda}{\pi w_0} \text{ for } z \gg R_z = \frac{\pi w_0^2}{\lambda} \tag{6.8}$$

where R_z is the Rayleigh range. Equation 6.8 shows that the beam's cross-sectional area at distance z from the waist is

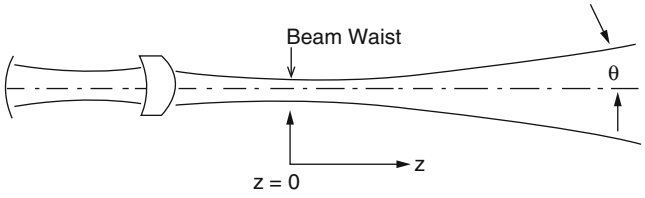


Fig. 6.4 Gaussian beam with external waist at $z = 0$

$$A = \pi[w(z)]^2 = \frac{\lambda^2 z^2}{\pi w_0^2} \tag{6.9}$$

From this we see that the beam's cross-sectional area can be reduced by selecting a shorter wavelength and by increasing the beam radius at the waist [21].

It is interesting to look at a few examples of the intensity of laser light. In a typical ruby laser, the concentration [22] of Cr^{+++} ions is about 2×10^{19} cm^{-3}, and population inversions are of the order of $3 \times 10^{16} cm^{-3}$. Crudely speaking, we can think of creating 3×10^{16} quanta/cm^3 in the lasing medium. Since we have arranged the laser so that the output is in a single direction and since photons move with the speed of light, we obtain $3 \times 10^{16} \times 3 \times 10^{10} = 9 \times 10^{26}$ quanta/cm^2 s from the laser. For ruby, the lasing wavelength is 6943 Å, and since the energy of each quanta is hv, one can readily calculate that the output is about $2.5 \times 10^8 W/cm^2$.

Let us compare this to the power that a hot body, say the sun, emits at the same wavelength with a similar bandwidth. This can be calculated by the use of Planck's radiation law knowing that $\Delta E = hv$, where h is Plank's constant and v is radiation frequency [23]. This is illustrated in Fig. 6.5 and is the case for any light source, whether laser, flame, incandescent body, etc.

$$U_\omega = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{\hbar\omega/kT}} \tag{6.10}$$

where U_ω is the energy, per unit volume and per unit bandwidth, radiated by a blackbody at temperature T, k is Boltzmann's constant, c is the velocity of light in vacuum, while $\omega = 2\pi v$ is angular velocity of photon and $\hbar = \frac{h}{2\pi}$ is reduced Planck's constant. The radiation leaves the blackbody source at rate c, so that the power radiated per unit area of the source and per unit bandwidth, is

$$I_\omega = \frac{cU_\omega}{4} = \frac{\hbar\omega^3}{\pi^2 c^2} \frac{1/4}{e^{\hbar\omega/kT} - 1} \tag{6.11}$$

If we use the sun's temperature of 6000 K, and $\lambda = 6943$ Å,

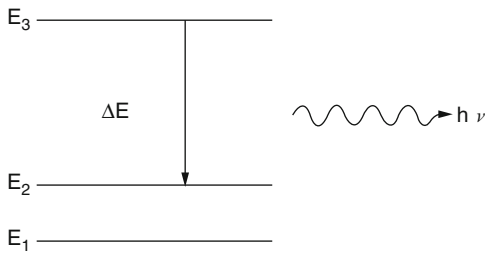


Fig. 6.5 Energy levels

$$I_\omega \approx 2 \times 10^{-5} \text{ erg/cm}^2$$

For the ruby laser, a typical line width is 3 Å, so $\Delta\omega \approx 1.2 \times 10^{12} s^{-1}$. Thus the power density at the source is

$$I \approx 2.5 \times 10^{12} \text{ erg/cm}^2\text{s} \approx 2.5 \text{ W/cm}^2$$

Thus the power density for comparable narrow bandwidth, nearly single-frequency light is much greater at a laser source than at a conventional hot body source, because laser light is coherent.

The propagation of laser light through the atmosphere poses a complex problem and it is discussed here. Suffice it to say that, as anyone who has driven on a foggy night certainly realizes, light is certainly scattered in the atmosphere. Lasers of high-power density pose even more difficult propagation problems because the high intensity warms the air and creates a density change across the beam. This variation in density refracts the light and causes beam spreading or "thermal blooming."

Consider briefly the propagation of laser light in free space or in vacuum. Under these ideal conditions, the only change in the power density is due to simple beam divergence. Since the typical laser emits light that is nearly unidirectional, the beam divergence is small. In fact, one feature of a laser is that the divergence is nearly at the diffraction limit, which is of the order of λ/a, where a is the diameter of the output aperture of the laser. For the ruby laser discussed above, this gives a divergence angle of [22]

$$\theta \approx \frac{6943 \times 10^{-8}}{1} \approx 7 \times 10^{-2} \text{ mrad}$$

for a 1-cm aperture. In practice, one needs to go to much trouble to realize this limit of divergence, but it has been done. More commonly, an "off-the-shelf" ruby laser might have a beam divergence of a few mrad.

The newcomer to lasers has usually heard about diffraction-limited beams and the consequent extreme directionality of laser light. He is usually surprised to discover that at long distances from the source, these beams have power densities that vary as the reciprocal of the square of the distance, like all radiating sources. To see this, consider a source of power P W, diameter A, and divergence angle θ, as shown in Fig. 6.6.

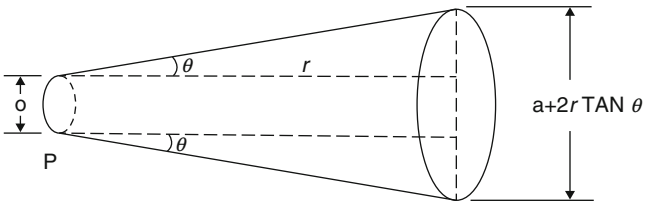


Fig. 6.6 Simplified sketch of laser beam divergence

At distance r from the source, the power density is

$$I = \frac{P}{\frac{\pi}{4}(a + 2r\tan\theta)^2}$$

or, since θ is very small and $\tan\theta \approx \theta$,

$$I = \frac{P}{\frac{\pi}{4}(a + 2r\theta)^2}$$

or

$$I = \frac{P}{\frac{\pi a^2}{4}\left(1 + \frac{2r}{a}\theta\right)^2} \tag{6.12}$$

From this expression, it is apparent that for large distance, such that $2r/a \gg 1$,

$$I = \frac{P}{\frac{\pi a^2}{4}\frac{4r^2\theta^2}{a^2}}$$

or

$$I = \frac{P}{\pi r^2\theta^2}$$

Or, since $\theta \approx \lambda/a$

$$I = \frac{P}{\pi r^2}\frac{a^2}{\lambda^2}$$

For example, consider a 10-kW beam of 10.6-μm wavelength and 10-cm aperture at 1-mi (i.e., a high-power CO_2 laser):

$$I = \frac{10^4 \times 10^2}{\pi(5280 \times 12 \times 2.54)^2\left(10 \times 10^{-4}\right)^2}$$

or

$$I \approx 12\ \mathrm{W/cm^2}$$

From Eq. 6.12, if we substitute I_0, the power density at the source, for $P/(\pi a^2/4)$ and recall that $\theta \approx \lambda/a$,

$$I = I_0 \frac{1}{\left(1 + 2r\frac{\lambda}{a^2}\right)^2} \tag{6.13}$$

From this expression, we can see that if r is small, there is little change in power density emitted by the source. The distance at which this is true is referred to as "near field," and the fine details of the beam pattern, such as local variation in intensity, hot spots, etc., are preserved in the near field. It is apparent from Eq. 6.13 that this near-field distance will be limited to r such that $I \approx I_0$, or

$$\frac{2r\lambda}{a^2} \ll 1$$

or

$$r_{\text{near field}} \ll a^2/\lambda$$

For lasers with exceptionally good optics that have a Gaussian distribution of power density across the beam, the near-field pattern will persist for distances on the order of a (a^2/λ) [5].

As a final comment on power densities at distance from laser sources, let us use Eq. 6.13 to calculate the distance at which the power density is halved: [22]

$$\frac{I}{I_0} = \frac{1}{2} = \frac{1}{\left(1 + \frac{2r\lambda}{a^2}\right)^2}$$

and

$$r_{1/2} = \frac{a^2}{2\lambda}\left(\sqrt{2} - 1\right)$$

For our illustration of a CO_2 laser with a 10-cm aperture, $r \approx 680$ ft or a little more than 0.1 mi. These calculations were done by J. T. Schriempf [22] and were reproduced here again.

6.5 Thermal Blooming Effects

High-energy laser propagation in the atmosphere requires consideration of self-induced beam expansion due to thermal blooming and random distortion due to atmospheric turbulence. The thermal blooming is a result of interaction between the laser radiation and the propagation path. A small portion of the laser energy is absorbed by the atmosphere. This energy heats the air causing it to expand and form a distributed thermal lens along the path. The refractive index of the medium is

decreased in the region of the beam where heating is the greatest, causing the beam to spread. Atmospheric turbulence is caused by random naturally occurring temperature gradients in the atmosphere.

Researchers are focusing on the design of beam control system for high-energy lasers. In particular, some folks are concentrating to compare traditional phase conjugation and open-loop techniques to a model-based optimal correction technique which modified the laser power and focal length. For light thermal blooming, phase conjugation is seen to be a reasonable control strategy. However, as the level of thermal blooming increases, phase conjugation performs increasingly worse. For moderate to heavy thermal blooming scenarios, the new technique is shown to increase peak intensity on target up to 50 % more than traditional compensation methods. Considering a ground-based continuous-wave laser operation in an environment with wind shows that in addition, the optimal correction technique is insensitive to errors in the model parameters. Assumption is that a tracking system provides target position and velocity. A reflection of the laser wave front off the target is useful but not required.

In case of ground-based high-energy laser (GBL), a typical platform is consistent of the following elements which (i.e., the mirrors and separators), in order of an outgoing wave, are described below and the optical path is depicted in Fig. 6.7:

1. **Deformable mirror**: Deforms the wave front taking into account the wave front received at the incoming and outgoing wave-front sensors.
2. **Beam splitter**: Allows a small amount of the laser to be fed to the wave-front sensor while reflecting the rest onto the deformable mirror.
3. **Outgoing wave-front sensor**: Detects the wave-front error before the laser is reflected off any mirrors.
4. **Turning mirror**: Reflects the beam.
5. **Tilt mirror**: High-bandwidth mirror which can point the beam in any direction, used to remove tilt errors from the wave front.
6. **Beam expansion**: Consists of a small convex mirror and a large concave mirror. It allows beam steering and focusing.
7. **Large turning mirror (traverse)**: Used for course pointing in combination with rotation of the whole beam expander. It has limited orthogonal motion capability creating a traverse axis for better dynamic performance.

As we said before, the purpose of a high-energy laser beam system is to deliver maxim power to a target. Several atmospheric effects decrease the effectiveness of such mission for the given system. These effects include both linear and nonlinear terms (see Sect. 6.3 in above). Diffraction, turbulence, jitter, and wander linearly decrease the intensity on target. If the nonlinear effects are ignored, any increase in intensity on target can be accomplished by increasing the laser power. When the nonlinear effects of thermal blooming is included, increasing laser power will not always be beneficial and can even reduce the level of transmitted power. Figure 6.4 shows the performance of an open-loop system by determining intensity on target as a function of laser power with and without blooming. As seen in Fig. 6.8, it is clear that thermal blooming must be considered when evaluating this system if the

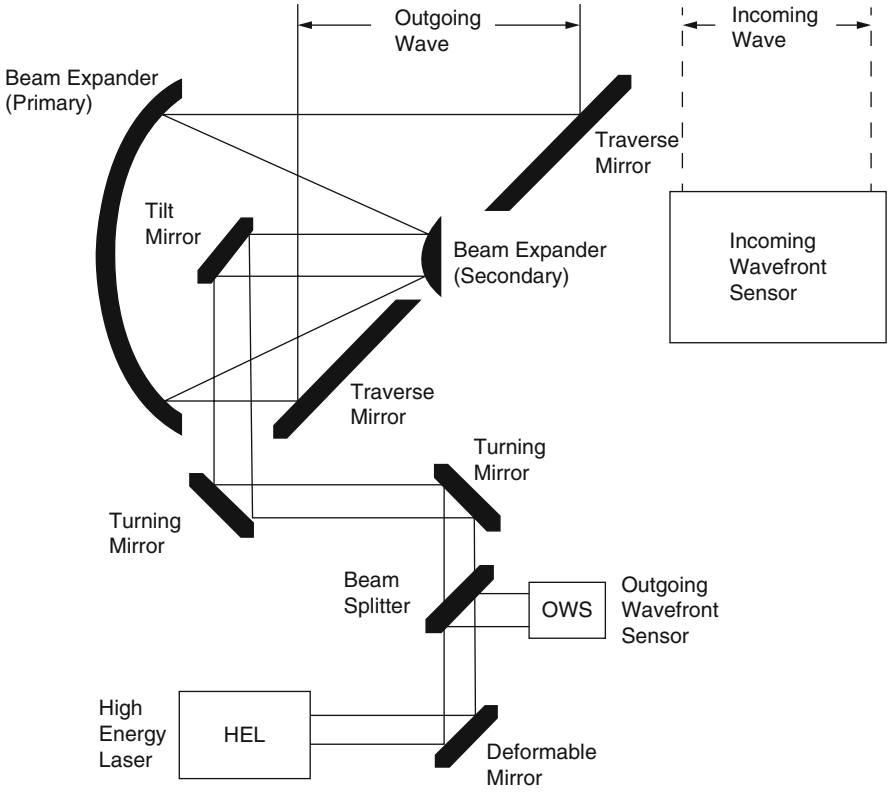


Fig. 6.7 Typical high-energy ground-based laser platform

laser is operating above 25×10^3 W. If thermal blooming is ignored in the design stage, the actual intensity on target will be a small fraction of what is expected.

Distortion caused by atmospheric turbulence is independent of the applied phase, and phase conjugation has the possibility of significantly reducing wavefront distortion at the target. However, the atmospheric distortion caused by thermal blooming is a function of the applied phase. It is shown that phase conjugation techniques are not optimal, and model-based controllers can improve the performance of directed energy systems. It has been recognized for some time that phase conjugation methods are prone to instability. The phenomenon of phase compensation instability (PCI) has been studied extensively [24, 25].

It is worth noting that thermal blooming produces a result that seemingly contradicts logic beyond a certain point, increasing the laser power can decrease the amount of laser energy delivering to target. This happens because the thermal distortions caused by the laser beam grow faster than the laser power. The more laser energy poured into the beam, the larger the percentage of energy that is bent away from the target. It is nature's reminder to the weapon system designers that bigger is not always better, and it sets an ultimate limit to the size of battlefield

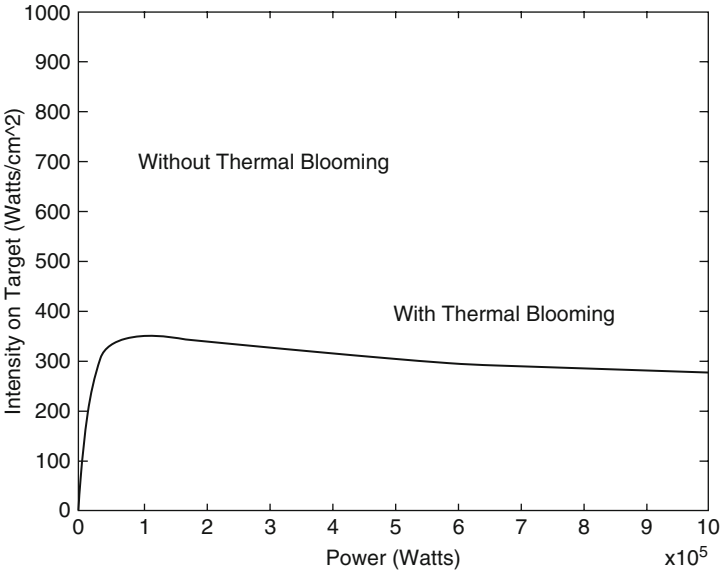


Fig. 6.8 Effect of thermal blooming on power-intensity curve

lasers. Note that this is not the case if the beam doesn't have to travel though atmosphere, so advocates of space-based laser weapons are free to propose laser as big as they like so long as they are not engaging in target during its boost phase from its launching platform. But the question is if they can build and deploy them to the orbit of their mission.

6.5.1 Mathematical Foundation of Thermal Blooming

Thermal blooming of high-energy lasers is a beam-spreading effect that can significantly reduce the effectiveness of laser systems both as directed energy weapons and as remote powering devices. When a high-energy laser propagates through a medium, a portion of the laser energy is absorbed by the medium. This absorbed power heats the medium causing it to expand, changing its index of refraction [26].

Thermal blooming is classified by the form of heat transfer that balances the absorbed power. The three cases of thermal conditions are:

1. Natural convection and forced convection.
2. Thermal condition occurs where there is no relative motion between the beam and the medium and when no natural convective velocities are established. Natural convection results when the absorbed power causes gas heating, which establishes convection currents.

3. By far the most important continuous-wave case is convection forced by wind and beam slewing.

A derivation of simple scaling laws for thermal blooming by J. Edward Wall [26] is defining basic nomenclature for adaptive optics and thermal blooming. The classical "bending into the wind" shape of a thermally bloomed wave was pictured. Additionally, the important relationship between intensity and wave-front error is examined by him, and he has introduced the representation of two-dimensional wave fronts by Zernike polynomials.

Thermal blooming is a highly nonlinear phenomenon. A simple method for analyzing the effects of thermal blooming is to start with the general wave optics equations and find a perturbation solution. The following equations completely specify steady-state thermal blooming in the ray optics limit. Further information and mathematical analysis can be found in a book by Hugo Weichel [21] where he has described mathematical analysis of laser beam propagation in the atmosphere.

6.6 Adaptive Beaming and Imaging in Turbulent Atmosphere

In order to deal with optical distortion that is caused by atmospheric turbulence and thermal blooming effects—corrective optics measurement, such as adaptive optical systems is required. Due to dynamic nature and behavior of these effects and their continuous changes of their conditions, the corrective optics would have to be adjustable accordingly, or in the jargon of the field, "adaptive." The idea is to use the optics to adjust the wave front leaving the laser beam weapon in a way that would compensate for the distortions that such beam encounters while going through atmosphere to engage with its designated target.

The extensive use of optical technologies for solving problems of information transfer, in form of a narrow-directional electromagnetic energy transport, and image formation in an outdoor atmosphere calls for the development of adaptive optical correction methods and devices. These methods call for an effective means of controlling the decrease in the efficiency of atmospheric optical systems caused by inhomogeneities in large-scale refractive indexes. These inhomogeneities are due to the turbulent mixing of atmospheric air masses and molecular and aerosol absorption in the channel of optical radiation propagation [23].

Ordinary transmissive optics is out of the question. The lens of the human eye is deformable or in a sense "adoptive," but essentially all other transparent materials used as optical lenses are rigid. Solid transparent media can be damaged by high-power laser beams because they, like the air, absorb a small fraction of the optical energy they transmit. There is some research into the possibility of developing gaseous optics to make the required corrections, but it is far from clear if that concept can be made practical for the laser weaponry [2].

Wide practical application of adaptive optical systems (AOS) has revealed a number of problems that call for the development of a theory of optical wave propagation under adaptive control conditions. A search for answers to these problems necessitates the development of detailed and adequate mathematical AOS models and the application of research methods such as numerical experiments that solve a system of differential equations describing optical wave propagation in the atmosphere.

Adaptive optical systems (AOS) that operate in real time allow one to:

- Improve laser radiation focusing on a target and hence increase the radiation intensity within the focal spot.
- Decrease the image blooming of astronomical and other objects m telescopes, increase image sharpness, and decrease the probability of object recognition errors.
- Decrease the noise level and increase the data rate in optical communication systems.

A wide variety of research is going on to investigate using numerical experiments (models). Numerical experiments allow the maximum number of parameters to be considered to correctly model AOS and to investigate practically any significant radiative characteristic—the effective size of the light spot, the peak radiant intensity, the radiation power incident on the receiving aperture, the statistical characteristics of the radiant intensity, and phase—in the context of a universal approach. A numerical experiment with applications to AOS allows one to predict the efficiency of various system configurations. Much time and considerable expense would be required to perform field experiment.

A method for numerical solution can be applied to two tasks:

1. High-power laser beam propagation in homogeneous media with absorption, and optical wave propagation through a random inhomogeneous turbulent atmosphere.
2. As high-power coherent laser beams propagate through a nonturbulent atmosphere, *thermal blooming* is one of the main factors causing distortion, along with turbulent fluctuations of the refractive index.

This nonlinear effect has the lowest energy threshold and arises as a result of absorption of part of the beam energy and the formation of the thermal inhomogeneities in the beam channel.

The longitudinal scale of variability for thermal inhomogeneities induced in the propagation channel of a high-power laser beam (thermal blooming) is comparable to the diffraction length of the beam. In the interval Δz, the equation for the phase screen can be approximated by the product of step length Δz and the refractive index distribution at the center of the interval $[z_l, z_l + \Delta z]$ [27]. This is based on assumption of corresponding to propagation of a wave from the plane z_l.

$$\varphi_l(\vec{\rho}) = k\Delta z \delta n\left(\vec{\rho}, z_l + \frac{1}{2}\Delta z\right) + O\left(\Delta z^2\right) \tag{6.14}$$

where

$k = 2\pi/\lambda$ is wave number

$\vec{\rho}(x, y)$ is the vector of coordinates in the beam cross section (the beam is directed along Oz axis)

$\delta n = (n - 1) \ll 1$ is the derivations of the refractive index from unity for $n_0 \approx 1$, with n_0 being vacuum refractive index

From Eq. 6.14, it follows that we have only to determine perturbations of the refractive index in some planes, the position of which are determined by the scheme of the splitting algorithm [27].

Heating of the medium that is caused by absorption of radiation energy includes variation of its density, which leads to a decrease in the refractive index related to the density ρ by the following law [28]:

$$\delta n = K\rho \tag{6.15}$$

where K is a constant equal to two-thirds of the polarization factor of a molecule or gas atom.

In the isobaric approximation, the density of the medium is explicitly related to temperature by the ideal gas law, so variations of the refractive index can be expressed through temperature variations [27]:

$$\delta n \approx \frac{\partial n}{\partial T}(T - T_0) = n_T^{'}\delta T \tag{6.16}$$

T_0 and T are corresponding to initial and final temperature, respectively. The isobaric approximation is valid for the normal atmospheric conditions. Exceptions that should be taken under consideration are, fast scanning of a continuous-wave (CW) high-power beam when the beam speed with respect to the medium is greater than the sonic speed and when the pulse duration τ_p is comparable with the acoustic time τ_s:

$$\tau_\mathrm{p} = \tau_\mathrm{s} = a/c_\mathrm{s} \tag{6.17}$$

where a is the beam size and c_s is the sonic speed.

When the isobaric approximation is valid, the distribution of the refractive index in the beam cross section is determined by the heat balance, which is described by the heat transfer equation for the temperature field $T(x, y, z)$: [27]

$$\frac{\partial T(x,y,z)}{\partial t} + \vec{V}_{\perp}\nabla T - \chi\Delta_{\perp} = \frac{\alpha}{\rho_0 C_{\mathrm{p}}} I \tag{6.18}$$

where

$\vec{V}_{\perp} = \left(V_x, V_y\right)$ is the transverse component of the beam velocity relative to the medium
χ is heat conductivity
ρ_0 is the specific density of the medium
α is the absorption coefficient
C_{p} is the specific heat at constant pressure

Vladimir P. Lukin and Boris V. Fortes [27] have provided a solution for the above equation and have argued when the isobaric approximation becomes invalid.

By creating an AOS that is smart enough with its onboard computer software system, we can overcome thermal blooming effects for better laser weaponry system particularly dealing with ground-based laser (GBL), airborne laser (ABL), and space-borne laser (SBL) in case of engagement with a target on ground (i.e., boost phase).

The most promising solution is to use a mirror in which the reflective surface is deformable, sometimes called a "rubber" mirror by engineers, although it really doesn't contain rubber. There are three basic types being developed. One is the segmented mirror in which there are many discrete segments, each of which is moved back and forth mechanically by a separate pistonlike device called an "actuator." Another is a mirror in which a reflective coating has been laid down on a base material that changes its shape when signals are applied to it. (In practice, the base material is generally a piezoelectric material, which changes size when an electrical voltage is applied to it. An array of electrodes applies different voltages to different parts of the mirror base [2] (Fig. 6.9).)

A third concept is similar in that the mirror has a continuous flexible surface, but in this case, the precise contour of the surface is shaped by an array of individually controlled mechanical actuators lying beneath it [30]. All these concepts have been demonstrated, although not in the sizes necessary for practical weapon systems.

Performance requirements are stringent for high-energy laser applications. The mirror must be able to withstand the laser's high power, a need often met by forcing a liquid coolant through holes in the body of the mirror. Getting fine enough control over the wave front of the beam requires many separate and ultraprecise control elements in the mirror. A mirror 16 cm (6.5 in.) in diameter requires at least 60 separate actuators, and proportionately higher numbers are needed for larger mirrors. The 16-cm mirror, together with its mount, should hit the scales at about 1000 kg (2200 Ib), and the 60 actuators should weigh no more than 800 kg (nearly 1800 Ib).

The shape of the optical surface must be precisely controlled. The mirror surface should be able to move back and forth over a total range of at least four times the laser wavelength. When the surface control is operating, the surface should be

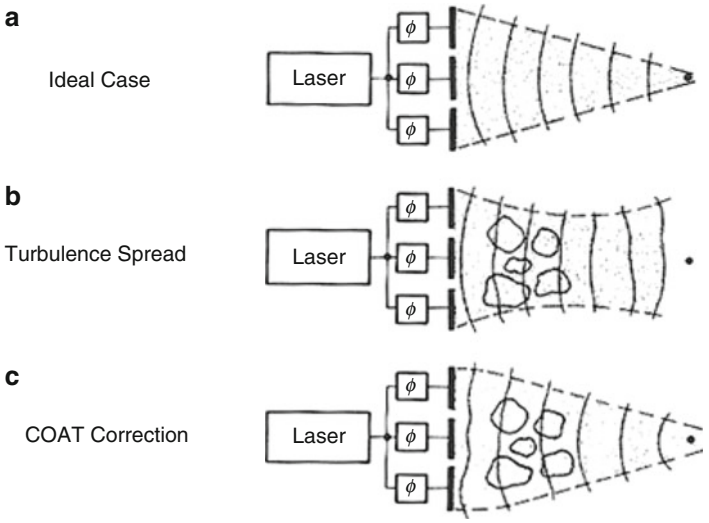


Fig. 6.9 The atmospheric transmission problem, and how it can be overcome—at least in theory. In a vacuum a laser beam could be focused tightly onto a small spot (*top picture* (**a**)). In the atmosphere turbulence, thermal blooming, and other effects spread out the beam over a much larger area (*middle* (**b**)). By using adaptive optics to adjust the wave front of the light emitted by the high-energy laser, the spreading can be reduced (*bottom* (**c**)). In these drawing, ϕ, indicates control systems that adjust the shape of the output mirror to control the laser wave front; they are compensating for atmospheric effects only in the bottom drawing [29]

within one-twentieth of the laser wavelength of the ideal shape. As if that isn't enough, adjustments in mirror shape have to be made about 1000 times a second to keep up with fluctuations in the atmosphere. Because the optical tolerances depend directly on laser wavelength, they get tighter at shorter wavelengths. This helps to offset the advantage of being able to use smaller optics at such wavelengths (Fig. 6.10).

Adaptive optics can help compensate for effects other than atmospheric distortions that might defocus a laser beam. Some turbulence is inevitable in the laser itself, as gases flow rapidly through the laser cavity and react to release energy. Corrections applied through deformable mirrors can help in precisely tracking targets and in finely focusing the laser beam onto a distant target, although gross mechanical motion of the mirror would be needed to provide full compensation for anything beyond small movements [2].

6.6.1 Adaptive Optics

No matter how powerful a laser is, it will never reach its target without optical components. The optical components not only "direct" the beam through the laser to its target, but they also relay the laser energy and, when required, correct for any

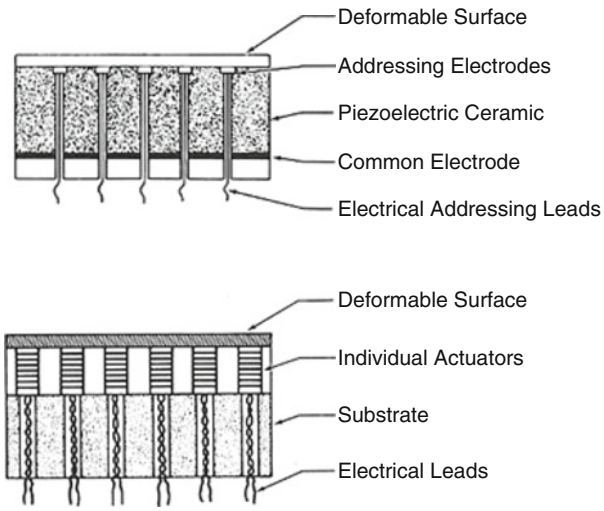


Fig. 6.10 Two types of active mirrors shown in cross section to indicate how the surface is controlled. In the mirror at top, the flexible surface layer rests on a block of piezoelectric ceramic, which changes its height when an electrical voltage is applied across it. Applying different voltages across different parts of the mirror alters its shape because the height of the piezoelectric material changes unevenly across the surface. In the mirror at bottom, the flexible surface layer covers an array of pistonlike actuators, which move back and forth in response to electrical signals, thus changing the shape of the mirror [29]

atmospheric turbulence that will distort the beam. The tremendous advances in optics have played a key role in convincing the Air Force that laser weapon systems can be produced. Without these successes by government laboratories and industry, high-energy laser weapons would be impossible.

The reason stars twinkle in the night sky is due to atmospheric turbulence, which also will distort and degrade any laser. This effect has especially severe effects for the shorter wavelength lasers, such as COIL. These systems require sophisticated optics in order to "pre-compensate" the laser beam for atmospheric turbulence 60 to pre-shape the laser beam; an adaptive optics technique is used. Over the past several years, the Air Force Research Laboratory, Phillips Research Site, and the Massachusetts Institute of Technology's Lincoln Laboratory have made significant strides in adaptive optics.

The principle behind adaptive optics is to use a deformable mirror to compensate for the distortion caused by the atmosphere. The system first sends out an artificial "star" created by a low-power laser. When that laser beam is scattered by the atmosphere, the scattering radiation is reflected back and measured so that the system knows just how much the atmosphere is distorting the laser. By feeding this information into a complex control system, the deformable mirror, with its hundreds of small actuators positioned behind the mirror, alters the surface of the mirror to compensate for atmospheric distortion. Thus, a high-energy laser can be "pre-distorted" so it will regain its coherence as it passes through the atmosphere.

The Starfire Optical Range at the Phillips Research Site has successfully demonstrated the adaptive optics technique. It has a telescope with the primary mirror made of a lightweight honeycomb sandwich, which is polished to a precision of 21 nm or approximately 3000 times thinner than a human hair. To compensate for the distortion caused by gravity, the primary mirror has 56 computer-controlled actuators behind its front surface to maintain the surface figure. The 3.5-m telescope adaptive optics system has a 941-actuator-deformable mirror that is controlled by a complex computer system. What has been accomplished at the Starfire Optical Range represents possibly the most significant revolution in optical technology in the past 10 years.

6.6.2 Deformable Mirror

Deformable mirror (DM) represents the most convenient tool for wave-front control and correction of optical aberrations. Deformable mirrors are used in combination with wave-front sensors and real-time control systems in adaptive optics. They are also finding a new use in femtosecond pulse shaping. The shape of the DM can be controlled with a speed that is appropriate for compensation of dynamic aberrations present in the optical system. In practice the DM shape should be changed much faster than the process to be corrected, as the correction process, even for a static aberration, may take several iterations. A DM usually has many degrees of freedom. Typically, these degrees of freedom are associated with the mechanical actuators, and it can be roughly taken that one actuator corresponds to one degree of freedom.

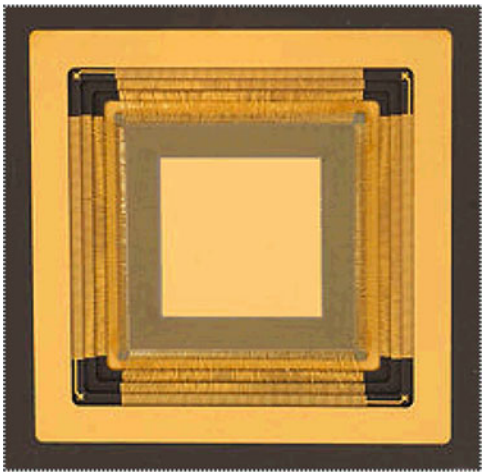
6.6.2.1 Deformable Mirror Concepts

Segmented concept: Mirrors are formed by independent flat mirror segments. Each segment can move a small distance back and forth to approximate the average value of the wave front over the patch area. Normally these mirrors have little or zero cross-talk between actuators. Stepwise approximation works poorly for smooth continuous-wave fronts. Sharp edges of the segments and gaps between the segments contribute to light scattering, limiting the applications to those not sensitive to scattered light. Considerable improvement of the performance of the segmented mirror can be achieved by introduction of three degrees of freedom per segment: piston, tip, and tilt. These mirrors require three times more actuators than piston-segmented mirrors, and they suffer from diffraction on the segment edges. This concept was used for fabrication of large segmented primary mirrors for the Keck telescopes.

Continuous faceplate concept: Mirrors with discrete actuators are formed by the front surface of a thin deformable membrane. The shape of the plate is controlled by

a number of discrete actuators that are fixed to its back side. The shape of the mirror depends on the combination of forces applied to the faceplate, boundary conditions (the way the plate is fixed to the mirror), and the geometry and the material of the plate. These mirrors are often the most desirable implementation, as they allow smooth wave-front control with very large—up to several thousand—degrees of freedom.

MEMS deformable: Mirror with 1020 actuators, from Boston Micromachines Corporation. MEMS concept mirrors are fabricated using bulk and surface micromachining technologies. MEMS mirrors have a great potential to be cheap. They can break the high price threshold of conventional adaptive optics. MEMS mirrors typically have high response rates, have high precision, and have no hysteresis, unlike other types of deformable mirrors. Boston Micromachines Corporation is one example of a company that produces MEMS Deformable mirrors.



Membrane concept: Mirrors are formed by a thin conductive and reflective membrane stretched over a solid flat frame. The membrane can be deformed electrostatically by applying control voltages to electrostatic electrode actuators that can be positioned under or over the membrane. If there are any electrodes positioned over the membrane, they are transparent. It is possible to operate the mirror with only one group of electrodes positioned under the mirror. In this case, a bias voltage is applied to all electrodes, to make the membrane initially spherical. The membrane can move back and forth with respect to the reference sphere.

Bimorph concept: Mirrors are formed by two or more layers of different materials. One or more of (active) layers are fabricated from a piezoelectric or electrostrictive material. Electrode structure is patterned on the active layer to facilitate local response. The mirror is deformed when a voltage is applied to one or more of its electrodes, causing them to extend laterally, which results in local mirror curvature. Bimorph mirrors are rarely made with more than 100 electrodes.

Ferrofluid concept: Mirrors are liquid-deformable mirrors made with a suspension of small (about 10 nm in diameter) ferromagnetic nanoparticles dispersed in a liquid carrier. In the presence of an external magnetic field, the ferromagnetic particles align with the field, the liquid becomes magnetized, and its surface acquires a shape governed by the equilibrium between the magnetic, gravitational, and surface tension forces. Using proper magnetic field geometries, any desired shape can be produced at the surface of the ferrofluid. This new concept offers a potential alternative for low-cost, high-stroke, and large number of actuator-deformable mirrors.

6.6.2.2 Deformable Mirror Parameters

Number of actuators: Determines the number of degrees of freedom (wave-front inflections) the mirror can correct. It is very common to compare an arbitrary DM to an ideal device that can perfectly reproduce wave-front modes in the form of Zernike polynomials. For predefined statistics of aberrations, a deformable mirror with M actuators can be equivalent to an ideal Zernike corrector with N (usually $N < M$) degrees of freedom. For correction of the atmospheric turbulence, elimination of low-order Zernike terms usually results in significant improvement of the image quality, while further correction of the higher-order terms introduces less significant improvements. For strong and rapid wave-front error fluctuations such as shocks and wake turbulence typically encountered in high-speed aerodynamic flow fields, the number of actuators, actuator pitch, and stroke determines the maximum wave-front gradients that can be compensated for.

Actuator pitch: Is the distance between actuator centers. Deformable mirrors with large actuator pitch and large number of actuators are bulky and expensive.

Actuator stroke: Is the maximum possible actuator displacement, typically in positive or negative excursions from some central null position. Stroke typically ranges from ± 1 to ± 10 μm. Free actuator stroke limits the maximum amplitude of the corrected wave front, while the inter-actuator stroke limits the maximum amplitude and gradients of correctable higher-order aberrations.

Influence function: Is the characteristic shape corresponding to the mirror response to the action of a single actuator. Different types of deformable mirrors have different influence functions; moreover the influence functions can be different for different actuators of the same mirror. Influence function that covers the whole mirror surface is called a "modal" function, while localized response is called "zonal."

Actuator coupling: Shows how much the movement of one actuator will displace its neighbors. All "modal" mirrors have large cross-coupling, which in fact is good as it secures the high quality of correction of smooth low-order optical aberrations that usually have the highest statistical weight.

Response time: Shows how quickly the mirror will react to the control signal and can vary from microseconds (MEMS mirrors) to tens of seconds for thermally controlled DM's.

Hysteresis and creep: Are nonlinear actuation effects that decrease the precision of the response of the deformable mirror. For different concepts, the hysteresis can vary from zero (electrostatically actuated mirrors) to tens of percent for mirrors with piezoelectric actuators. Hysteresis is a residual positional error from previous actuator position commands and limits the mirror ability to work in a feed forward mode, outside of a feedback loop.

6.6.3 Large Optical Systems

In addition to adaptive optics, large mirrors, either on the ground or in space, are needed to expand and project the laser energy onto the missile. Several significant large optics programs were conducted in the late 1980s and early 1990s. The Large Optics Demonstration Experiment (LODE) established the ability to measure and corrected the outgoing wave front of high-energy lasers. The Large Advanced Mirror Program (LAMP) designed and fabricated a 4-m diameter lightweight, segmented mirror. This mirror consists of seven separate segments that are connected to a common bulkhead. The advantages of building a mirror in segments are to reduce the overall weight and fabricate larger mirrors. In addition, each segment can be repositioned with small actuator motors to slightly adjust the surface of the mirror. The program's finished mirror successfully achieved the required optical figure and surface quality for a space-based laser application.

6.6.4 What Is Phase Conjugation in Optics?

Phase conjugation is a fascinating phenomenon with very unusual characteristics and properties. It operates somewhat like holography, but it is a dynamic hologram, whose "holographic plate" is defined by interfering wave fronts in a nonlinear optical medium, rather than etched as a static pattern on a glass plate. In this section, it is provided an intuitive explanation of the essential principles behind phase conjugation (Fig. 6.11) [31].

Let us begin with the properties of a phase conjugate mirror. A phase conjugate mirror is like a mirror, in that it reflects incident light back toward where it came from, but it does so in a different way than a regular mirror.

In a regular mirror, light that strikes the mirror normal to its surface is reflected straight back the way it came (A). This is also true of a phase conjugate mirror (B). When the light strikes, a normal mirror at an angle, it reflects back in the opposite direction, such that the angle of incidence is equal to the angle of reflection (C).

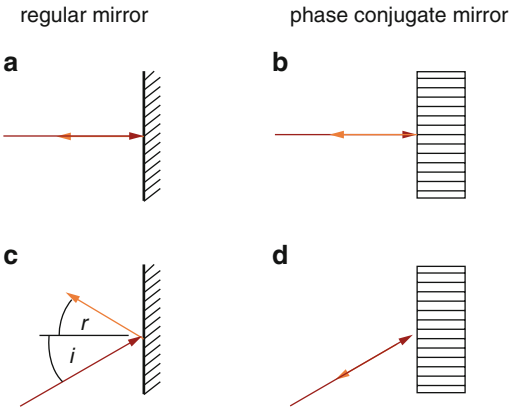


Fig. 6.11 Image of regular and phase conjugate mirror

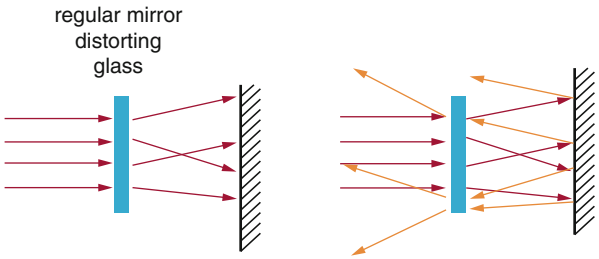


Fig. 6.12 Image of regular mirror

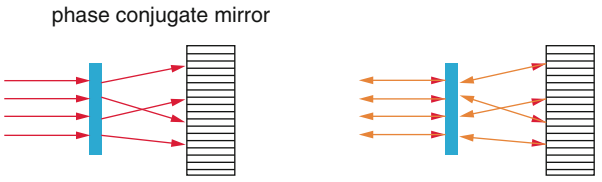


Fig. 6.13 Image of phase conjugate mirror

In a phase conjugate mirror, on the other hand, light is always reflected straight back the way it came from, no matter what the angle of incidence (D)

This difference in the manner of reflection has significant consequences. For example, if we place an irregular distorting glass in the path of a beam of light, the parallel rays get bent in random directions, and after reflection from a normal mirror, each ray of light is bent even farther, and the beam is scattered (Fig. 6.12).

With a phase conjugate mirror, on the other hand, each ray is reflected back in the direction it came from. This reflected conjugate wave therefore propagates backward through the distorting medium, and essentially "un-does" the distortion, and returns to a coherent beam of parallel rays traveling in the opposite direction (Fig. 6.13).

6.6.4.1 How Does the Phase Conjugate Mirror Do That?

In linear optics, light waves pass through each other transparently, as if the other waves were not there, and the same is true of the ripples on a pond that also pass through each other totally unaffected after they cross. But almost any optical, or other wave phenomenon, will go nonlinear if the amplitude is sufficiently high, and that is also true of water waves, to help our intuition. When waves in a ripple tank are driven too strongly, they lose their perfect sinusoidal shape and form sharper peaks between wide valleys, like wind-driven waves on the ocean. A most extreme nonlinear wave is seen in breaking waves on the beach, whose towering crests carry with them a slug of moving water. Waves of this sort do not pass through each other transparently, but they collide and rebound energetically like colliding billiard balls. In reality, nonlinear waves exhibit both linear and nonlinear components, so that colliding waves will simultaneously pass mostly through each other unaffected, and at the same time, some portion of those waves collide with and rebound off each other, creating reflections in both directions. This concept of waves colliding and rebounding provides the key insight into understanding the otherwise mysterious phenomenon of phase conjugation. This antiparallel rebounding of a ray of light in nonlinear optics, along with *Huygens's Principle* of wave propagation, is sufficient to explain some of the bizarre time-reversed reconstruction principles in phase conjugation, which is the principle that mirrors an observed property of perceptual reification.

6.6.4.2 Huygens's Principle

Huygens's principle states that a wave front is mathematically equivalent to a line of point sources all along that front, because the outward-radiating rays from adjacent point sources along the front eventually cancel by destructive interference, leaving only the component traveling in a direction normal to the local orientation of the front. This principle has an interesting spatial consequence that if the flame front has a shape, whether curved convex or concave or a zigzag or wavy line pattern, the shape of that wave front has a profound influence on the pattern of propagation of that front.

6.6.4.3 Two-Wave Mixing

The interactions between nonlinear waves is illustrated by the phenomenon of two-wave mixing, performed by projecting two laser beams to cross in the volume of a nonlinear optical medium. Figure 6.14a shows two laser beams, B_1 and B_2, that intersect through some volumetric region of space. In the volume of their zone of intersection, a pattern of standing waves emerges in the form of parallel planes, oriented parallel to the bisector of the angle between the two beams, as shown in

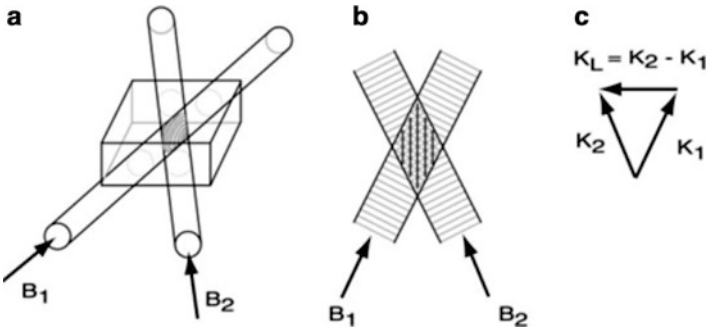


Fig. 6.14 (**a**) Two laser beams B1 and B2 that cross, create an interference pattern in their zone of intersection. (**b**) Constructive interference creates a pattern of high amplitude in parallel planes, parallel to the angular bisector of the two beams, with planes of low amplitude in between. (**c**) The wave vector diagram for the crossing beams, including a new lattice vector K_L that corresponds to the difference vector between the crossing beams

Fig. 6.14a. Figure 6.14b shows in two dimensions how the wave fronts from the two beams intersect to produce high amplitude by constructive interference along the vertical lines in the figure, interleaved with planar nodes of low or zero amplitude in between, due to destructive interference.

In linear optics, this interference pattern is a transient phenomenon that has no effect on anything else. However if the crossing of laser beams occurs in the transparent volume of a nonlinear optical medium, as suggested by the rectangular block in Fig. 6.14a, and if the amplitude of the beams is sufficiently large, the interference pattern will cause a change in the refractive index of the nonlinear medium in the shape of those same parallel planes, due to the optical Kerr effect. The alternating pattern of higher and lower refractive index in parallel planes behaves like a *Bragg diffractor*.

6.6.4.4 Bragg Diffraction

Bragg diffraction is distinct from regular diffraction by the fact that the diffracting element is not a two-dimensional grating of lines etched on a flat sheet, as in standard diffraction, but a solid volume containing parallel planes of alternating refractive index. Bragg diffraction was first observed in X-ray crystallography as a sharp peak of reflection at a particular angle of incidence to the crystal lattice planes. The crystal layers behave much like a stack of partially silvered mirrors, each plane passing most of the light straight through undiminished, but reflecting a portion of that light like a mirror, with the angle of reflection equal to the angle of incidence.

However because of interference between reflections from successive layers at different depths, Bragg diffraction is stronger at those angles of incidence that promote constructive interference between reflected rays but weakens or disappears

altogether at other angles where the various reflected beams cancel by destructive interference. Maximal diffraction occurs at angles that meet the Bragg condition, that is,

$$2d \sin \theta = n\lambda$$

where θ is the angle of the incident ray to the plane of the reflecting surface, d is the distance between adjacent planes, λ is the wavelength of light, and n is an integer. In words, Bragg reflection occurs at angles of reflection where the path length difference between reflections from adjacent planes differs by an integer number of wavelengths.

6.6.4.5 Reciprocal Lattice Wave Vector Representation

The phase matching constraint enforced by the Bragg condition can be seen most easily in a Fourier space called the reciprocal lattice representation. Each beam is represented by a wave vector whose direction is normal to the wave fronts of the corresponding beam and whose magnitude is proportional to the inverse of the wavelength or spacing between successive wave fronts of the beam. This is a Fourier representation in that the magnitude of the wave vectors is proportional to the frequency of the corresponding wave. Mathematically, the magnitude k of the wave vector of a wave of wavelength l is given by

$$k = 2\pi/\lambda$$

The convenience of this representation is that the wave vectors of waves that are phase matched so as to be in a mutually constructive relationship, form closed polygons in this space, and this can be used to determine whether the Bragg condition is met.

Figure 6.14c shows the wave vector representation for the crossing laser beams depicted in Fig. 6.14a. The wave vectors K_1 and K_2 are oriented parallel to their corresponding beams B_1 and B_2. The parallel planes of a Bragg diffractor, such as a crystal composed of parallel planes, can also be expressed as a wave vector because it behaves very much like a beam of coherent light to an incident beam that strikes it. As with wave vectors, the direction of this lattice vector K_L is normal to the planes of the grating, and the vector magnitude is proportional to the inverse of the spacing between lattice planes. Figure 6.15a shows the nonlinear optical element replaced by a functionally equivalent crystal with lattice planes parallel to those of the standing wave. The vector diagram of Fig. 6.14c shows the lattice vector K_L that would be required for the phase matching relation dictated by the Bragg condition to hold. In terms of wave vectors in the reciprocal lattice representation, the Bragg condition holds when

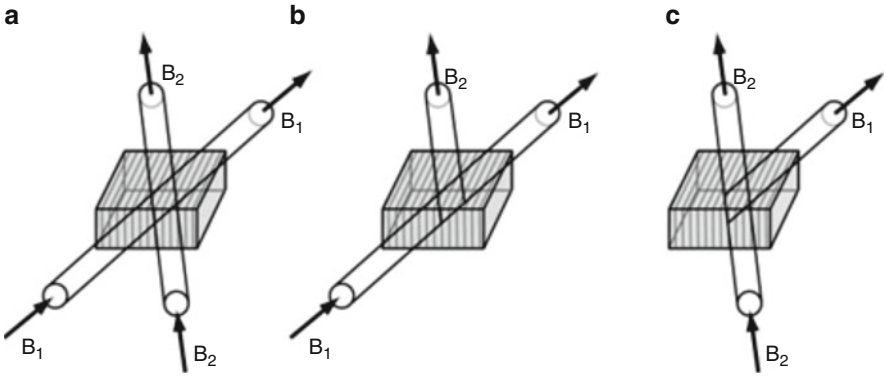


Fig. 6.15 (**a**) Nonlinear optical element replaced by functionally equivalent crystal with lattice planes parallel to the original standing waves and with the same spacing as the standing waves. (**b**) If beam B_2 were shut off, then beam B_1 together with the crystal would recreate B_2 by Bragg diffraction. (**c**) Conversely, if beam B_1 were shut off, then beam B_2 together with the crystal would recreate B_1 by Bragg diffraction

$$K_1 + K_L = K_2$$

or equivalently,

$$K_2 - K_L = K_1$$

The lattice vector acts in opposite directions on K_1 and K_2, which is why it is added to one but subtracted from the other. Note how the lattice vector is oriented normal to the planes of the lattice, which are parallel to the angular bisector of the two beams, as required for the angle of incidence to equal the angle of reflection. For example if the lattice spacing were somewhat larger than that dictated by the Bragg condition, that would make the lattice vector shorter, and the three vectors would no longer form a closed triangle, and thus little or no Bragg refraction would be expected to occur with that crystal, that is, the light would pass through with little or no reflection. Bragg refraction could be restored, however, by re-aligning either or both beams to make their wave vectors meet the shorter lattice vector.

6.6.4.6 Magical Reification

The magic of nonlinear optics is that when laser beams cross in the volume of a nonlinear optical medium, as depicted in Fig. 6.14a, the wave vector of the resultant nonlinear standing wave pattern automatically takes on the configuration required by the Bragg condition, no matter what the angle of intersection of the two beams. So although Bragg reflection occurs off a crystal only for certain specific angles that meet the Bragg condition, laser beams that cross in a nonlinear optical medium

create a standing wave whose lattice vector is automatically equal to the difference between the two crossing beams, or,

$$K_L - K_2 = K_1$$

This is a remarkable constructive or generative function of nonlinear optics, creating a whole new waveform out of whole cloth, equal to the difference between two parent wave forms. This magical act of creation can be understood as a property of the fundamental resonances in the nonlinear optical material set up by the passage of high-amplitude laser beams. The laser beam sets up a resonance in the electrons that are attached to the molecules in the optical material, that makes them vibrate in sympathy with the passing wave. The difference in nonlinear optics is that this resonance takes energy to establish, as if the electron had a certain momentum to be overcome or a capacitor that must absorb a certain charge, so that the optical material does not react instantaneously to the passing light, but with a certain energetic time lag, that borrows energy from the wave when the wave first turns on and repays that energy debt when the wave is shut off again, like a capacitor discharging through a resistor or a mass-and-spring system returning to center after wave passage. This is what makes nonlinear optics automatically balance the vector equation. If one wave vector deflects the electron this way and another deflects it that way, the electron needs to return back to the center before it can start the next cycle, and that returning back to the center is what closes the wave vector diagram.

If the pattern of standing waves were somehow frozen as a fixed pattern of alternating refractive index, as in a layered crystal, as suggested in Fig. 6.15a, then this crystal would behave like a hologram that can restore the pattern of light if one of the input beams is removed. For example, Fig. 6.15b shows beam B_1 refracted by the functionally equivalent crystal lattice to produce a reflected beam in the direction of the original beam B_2, and Fig. 6.16c shows beam B_2 refracted by the equivalent crystal lattice to recreate the original beam B_1. The reification in two-wave mixing has created a difference vector that has created a redundancy in the representation that allows either one of the input signals to be removed without loss of information.

If another analogy might be helpful, consider water flowing over sand, and creating little rippling dunes, and the rippling dunes in turn force the water to ripple over them, the flowing water and the rippling sand modulating each other by conforming to each other energetically. You can see the dunes eroding constantly from their flow-ward side, and building back up again on their leeward side, causing the little sand dunes to advance slowly to leeward, all in lock step with each other and with the corresponding ripples in the water. If you could instantly smooth the sand flat but preserve the rippling pattern in the water flow, it would immediately reestablish the ripples in the sand, by allowing the sand to accumulate in the stagnant parts of the flow. In fact, the rippling pattern would automatically reestablish itself naturally anyway, due to the fundamental dynamics of the water/sand interaction. Likewise, if the sand were frozen to a static plaster cast of the ripple pattern, that pattern would coerce any water flowing over it to conform to its

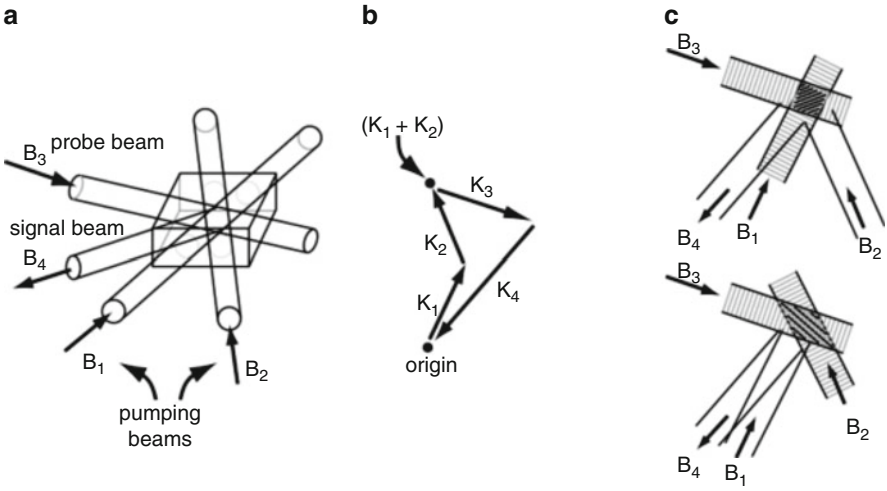


Fig. 6.16 (**a**) A third beam B_3 is directed into the intersection of the other beams, which produces a fourth beam B_4. The angle of that new beam can be calculated from the wave vector diagram as shown in (**b**). This can be seen intuitively as (**c**): an interference that forms between B_3 and B_1, followed by a reflection of B_2 by that pattern to create B_4, or alternatively it can be seen as (**d**) an interference between B_3 and B_2, followed by a reflection of B_1 by that interference pattern to create B_3

pattern of ripples, which the water would happily comply with, if the ripples are of the right natural frequency.

The nonlinear standing wave establishes an energy coupling between the two intersecting waves, such that one wave can "pump" or amplify the other. For example, if B_1 is of higher amplitude than B_2, then the interference pattern between B_1 and B_2 reflects some of the energy of B_1 in the direction of B_2, as in Fig. 6.15a, whereas if B_2 is of higher amplitude than B_1, some of the energy of B_2 is reflected in the direction of B_1, as in Fig. 6.15b. In fact, whether the two beams are of equal amplitude or not, some portion of B_1 is always lost to B_2 through the crystal, while some portion of B_2 is lost to B_1, as suggested in Fig. 6.15a, so the net energy transfer always flows from the higher amplitude beam toward the lower. That is, the two waves are intimately coupled through the nonlinear standing wave, energy-wise, and this energy coupling is what allows phase conjugation to produce an amplified reflection.

6.6.4.7 Degenerate Four-Wave Mixing

To create a phase conjugate mirror, we add a third probe beam, B_3, to intersect with the other two beams in the nonlinear optical element as shown in Fig. 6.16a. This creates a fourth signal beam B_4 which will eventually be our phase conjugate beam

after one last modification. This configuration is known as degenerate four-wave mixing. (The word degenerate refers to the fact that the frequencies of all four beams are equal, as required for the simplest form of phase conjugation exemplified here.) The direction of that fourth beam can be computed from the vector diagram shown in Fig. 6.16b, by the principle that the fourth beam will exactly cancel or balance the sum of the other three vectors, or,

$$K_1 + K_2 + K_3 + K_4 = 0$$

Again, this is dictated by the phase matching constraint, whereby the only waves that will emerge are those that reinforce each other constructively, and the reciprocal wave vector diagram helps identify the conditions under which that constraint is met. If the pumping beams K_1 and K_2 remain fixed, then whichever way the probe beam wave vector K_3 is directed from the point $(K_1 + K_2)$ in the vector diagram, the conjugate beam will always return back to the origin, as shown in Fig. 6.16b.

There are two ways that this phenomenon can be understood intuitively. We can say that probe beam B_3 interferes with pumping beam B_1 to produce an interference pattern as shown in Fig. 6.16c along their angular bisector, then beam B_2 reflects off that interference pattern to produce the signal beam B_4 (angle of reflection equals angle of incidence). Alternatively we can say that the probe beam B_3 interferes with other pumping beam B_2 to produce an interference pattern as shown in Fig. 6.16d, then beam B_1 reflects off that interference pattern to produce the signal beam B_4. It is more accurate however to think of all four beams as interlocked in a four-way energy coupling consummated by the newly created signal beam that appears so as to balance the vector equation and maintain phase coherence between all four beams. In other words, both interference patterns of Fig. 6.16c, d coexist simultaneously along with the original pattern of Fig. 6.14a, interlocking the four beams in a mutually interdependent energy relation.

6.6.4.8 Phase Conjugate Mirror

All we need to do to complete the phase conjugate mirror is to orient beams B_1 and B_2 antiparallel to each other, so that in vector terms $K_1 + K_2 = 0$. This in turn means that $K_3 + K_4 = 0$, which means that the reflected beam B_4 must be the phase conjugate of the probe beam B_3. Figure 6.17a shows the configuration required for phase conjugation. Pumping beams B_1 and B_2 are projected into the nonlinear optical element from opposite directions where they interfere to form a nonlinear standing wave. The probe beam B_3 can now be projected into the mirror from any direction, and this will produce the phase conjugate beam B_4 superimposed on B_3 but traveling in the opposite direction as a "time-reversed" reflection. The summation of B_3 and B_4 traveling in opposite directions converts the two waves into a standing wave that oscillates without propagation if they are of equal amplitude, otherwise there will be a net propagation in the direction of the higher amplitude beam. Figure 6.17b shows the wave vector diagram showing how if $K_1 + K_2 = 0$,

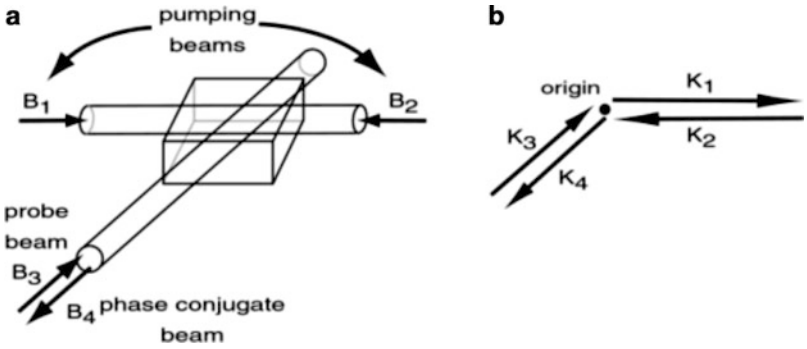


Fig. 6.17 (**a**) A phase conjugate mirror is produced by antiparallel pumping beams B_1 and B_2 that cross in opposite directions in the nonlinear optical element. When a third probe beam B_3 is projected into the mirror from any direction, a phase conjugate beam B_4 will appear as a time-reversed reflection of the probe beam in the direction from whence it came. (**b**) The wave vector diagram shows how if $K_1 + K_2 = 0$, then $K_3 + K_4$ also equals 0, and thus B_4 must be the phase conjugate of B_3

then $K_3 + K_4$ also equals 0 no matter what angle the probe beam enters the mirror, and thus B_4 must be the phase conjugate of B_3.

If the pumping beams are provided at high amplitude, then the energy built up in the nonlinear standing wave can spill over to the conjugate wave, creating an amplified reflection of the incoming wave back outward in the direction from whence it came. This is the phase conjugate mirror produced by degenerate four-wave mixing.

6.6.4.9 Phase Conjugation and Optical Black Magic

Adaptive optics is more than simply deformable mirrors. A control system is needed to determine how much to adjust the mirror's shape. Extensive theoretical work has been done on the propagation of high-energy laser beams through the atmosphere, but theory is not enough. Some of the most important distortions are caused by random atmospheric turbulence that theory cannot predict. The effects influencing light along the path the beam is going to travel must be measured, and that information must be converted into a control signal. This means that the control system must receive light returning along the beam path and analyze what has happened to it. This is by no means an easy process, and the details are well beyond the scope of this book [2].

After the control system has measured the effects that the beam will be subjected to, the type of compensation required must be determined. This process is called phase conjugation. It is a complex operation in which the measured effects of turbulence are used to create a laser wave front that will undo what the turbulence did, making it possible to produce a tight focal spot on the target. The precise method by which phase conjugation works is too complex to describe here; suffice

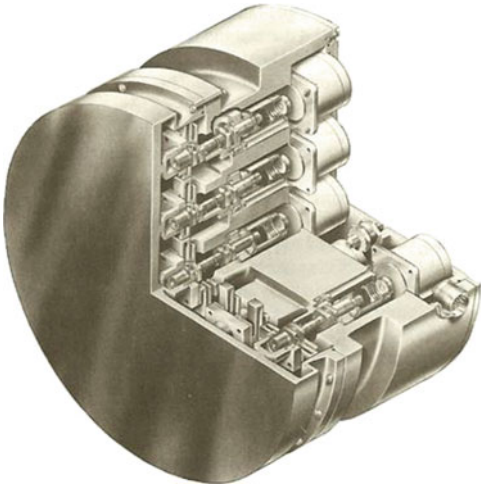


Fig. 6.18 Cutaway view of a 19-actuator-deformable mirror built by Rockwell International's Rocketdyne Division shows the complexity of adaptive optics. This mirror is 16 in. (40 cm) in diameter and weighs 100 lb (45 kg)

it to say that in theory, the technique can be used to compensate for aberrations inside the laser and in the atmosphere [21]. The critical corrections are made by adjusting the relative phase of different parts of the laser beam, that is, by making parts of the laser beam slightly out of step with each other, instead of staying in the normal lockstep of the light waves in a laser beam. Interestingly, changes in the intensity pattern of the laser beam are less important in compensating for atmospheric effects than the more subtle phase shift (Fig. 6.18) [2].

Beam distortions are not so serious in space, where there is no air to get in the way, but the tremendous distances involved to engage with target present. A massive amount of scientific manpower and defense money is going into development of adaptive optics. The topic is a common and hot one among scholar of this subject.

6.7 Target Effects

Once the beam reaches the target, it deposits part of its energy there. This involves a complex interaction between beam and target that depends strongly on the nature of the beam and property of the target and which ultimately determines how much of the energy in the beam is transferred to the target. Only after the energy is transferred to the target can it do any damage.

No one is seriously thinking of using a laser beam to completely vaporize any military targets. Instead, a continuous laser beam would cause physical damage by heating a target until the beam melted through the skin and lethally damaged some internal components. The actual type of damage would depend on the target and where it was illuminated. Drilling a hole in a fuel tank could cause an explosion.

Disabling the device called a "fuse," which triggers the explosion of a warhead, would prevent a bomb from exploding or alternatively might trigger a premature explosion of the warhead, in a place where it would not damage the intended target, but could cause considerable damage to other objects and people. Knocking out the control or guidance system could make a missile land far from its intended target [2].

A continuous laser beam can't do damage instantaneously. Heating the target to the required temperature would probably take a few seconds, depending on the laser power and the nature of the target; exact requirements are classified by the government but can be estimated doing few mathematical analysis. The illumination time is long enough for the beam to wander off the target spot and let the heated area cool off. Techniques called "countermeasures," described in many related defense papers, could be used to reduce the amount of energy that the beam could deposit on the target. Other types of interactions could also help protect the target [2].

There is considerable interest in substituting a rapid series of short laser pulses for a continuous beam. As mentioned earlier, this might simplify the task of getting the beam through the air to its target. The abrupt heating and cooling could cause thermal shock, sufficient to shatter materials such as glass. A short, intense pulse could also rapidly evaporate a burst of material from the surface, generating a shock wave that would travel through the target and could cause mechanical damage. (Evaporation caused by a continuous beam would be more gradual and would not cause a shock wave.) The combination of thermal and mechanical damage and heating effects caused by a series of short, closely timed laser pulses does a better job of breaking through sheet metal of target than either heating or laser produced shock waves can do by themselves [2].

Physical damage is not the only way a laser beam can disable a target. The beam could also attack sensors that guide weapons to their targets, blinding or disabling them by other means. Particle beams, microwaves, and X-rays have their own distinct ways of producing damages to target [2].

6.7.1 Measured Characteristic of Target Both Optically and Thermally

The purpose of measuring the characteristic of target is the absorptance of the target at the laser wavelength that is needed for model calculations of the target's thermal response. Ideally, the absorptance should be known as a function of temperature. However, if this is not available, the initial absorptance at ambient temperature is still a vital input to the thermal model and high-energy laser weaponry [32].

For accurate lethality estimates for a metal-skinned target, it is very desirable to know the target's absorptance at the laser wavelength and as a function of temperature. This can be done either with optical measurement approach that requires special equipment that is not generally available and is limited to power levels that

would not contaminate the hardware with sample ejecta. In lethality testing at high-power levels, it is possible to get absorptance values in real time using simple thermocouple instrumentation as discussed below [32].

6.7.2 Target Absorptance Optical Approach

For opaque (optically dense) targets as can be assumed for standard cases, the absorptance $A(\theta)$ at a given incidence angle θ can be inferred from a measurement of total hemispherical reflectance $R(\theta)$ at that incidence angle:

$$A(\theta) = 1 - R(\theta)$$

There are a number of government and commercial laboratories that perform spectrally resolved reflectance measurements. These give the *directional hemispherical reflectance* (DHR) of a sample as a function of wavelength over a range of interest. The sample is positioned in an integrating sphere and is illuminated at a fixed incidence angle by a spectrally resolved light source. This is due to the fact that absorptance of some materials is polarization dependent, and, therefore, the light source should either be unpolarized or measurements should be made with orthogonal polarizations. For more information on polarization subject, reader should refer to Appendix F this book. Commercial and custom reflectometers have the capability of heating the sample in situ to temperatures as high as 500 °C. Note that most laboratories restrict heating to temperature below the threshold for decomposition. Absolute measurement uncertainly is typically in the range of 0.01–0.02.

An alternative approach, illustrated in Fig. 6.19, has been pioneered by the Air Force Research Laboratory. The instrument, in what is known as the temperature-dependent reflectance of aerospace materials (TRAM) facility, is a hemi-ellipsoidal gold-coated dome. The sample is heated by a laser whose wavelength is different from the wavelength of interest provided by the probe beam or weaponry beam. The sample is located at one of the foci of the hemi-ellipsoid, and the entrance port of an integrating sphere is located at the other focal point.

In order for the integrating sphere detector to distinguish the probe beam signal from the heater beam, the probe beam is chopped at a relative high frequency (300 Hz) and a phase-sensitive ("lock-in") amplifier is employed. Additionally, a filter having high reflectance at the heater laser wavelength is positioned in front of the detector.

A pyrometer or thermographic imaging camera is employed to measure the sample's temperature as it is heated.

Note that for partially transparent materials, one must also measure their transmittance. Recent modifications to the TRAM instrument have permitted simultaneous measurements to be made of reflectance and transmittance as the sample is

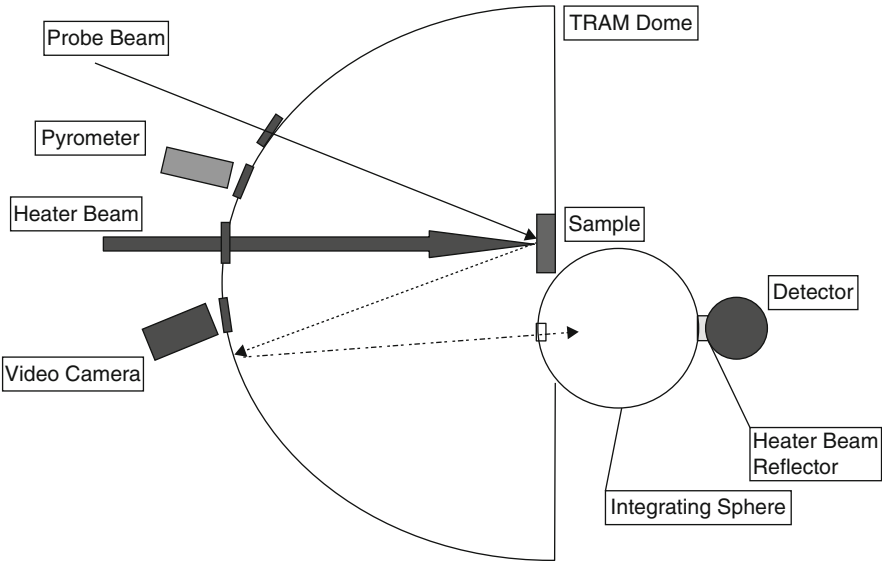


Fig. 6.19 Schematic of TRAM reflectometer

heated. However, this capability is still under development and not yet fully proven [32].

6.7.3 Target Absorptance Thermal Approach

The thermal response of a metal plate may be used to estimate the target's absorptance as a function of temperature. Typically, the temperature of the rear surface is measured with attached thermocouples and/or by use of a thermographic imaging camera. In the latter case, the rear surface is coated with a high emittance, thermally stable black paint.

Under conditions where two- or three-dimensional heat conduction effects can be ignored, a closed form analytical solution has been found, and it is shown in Sect. 6.6.4 below for the response of a finite thickness slab of material insulated in both faces to a constant uniform heat flux on one surface. This solution, which assumes constant thermal properties and no phase change, is given by

$$T(z,t) = T(z,0) + \frac{AIt}{\rho C_p L} + \frac{AIL}{k}\left\{\frac{3(1-z/L)^2-1}{6} - \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\exp\left(-\kappa n^2\pi^2 t/L^2\right)\cos\left(n\pi(1-z/L)\right)\right\} \tag{6.19}$$

where the axial coordinate in the original formula has been replaced by $L - z$ and the heat flux is assumed to be the absorbed irradiance at $z = 0$. Of course, from initial temperature to melt, themophysical properties of most materials vary considerably, but for limited temperature changes, this assumption is reasonable if average values are used. The limitations are further discussed below. Note in the above equation that A is the absorptance, I is irradiance (W/cm^2), t is time (s), ρ is density (g/cm^3), C_p is the specific heat (J/g K), L is the slab thickness (cm), k is the thermal conductivity (W/cm K), and κ is the thermal diffusivity (cm^2/s). The first term in the equation is the initial temperature, and the second term in the equation is just the linear rise in average plate temperature. The summation in the bracket is a strong function of the axial Fourier number, $f_{Na} = \kappa t/L^2$, the ratio of the duration of the irradiance to the characteristic conduction time through the thickness L. For $f_{Na} > 0.3$, the summation is approximately zero, and the first term in the bracket dominates, i.e., the temperature distribution develops a quasi-steady-state quadratic gradient superimposed on the average plate temperature. For small values of f_{Na}, the sum must be calculated to capture the transient; however, the sum converges rapidly with fewer than ten terms required for reasonable accuracy at $f_{Na} = 0.005$. The behavior of the bracketed term in Eq. 6.19 is shown in Fig. 6.20 below.

The thermal diffusivity of a relative low thermal conductivity material such as steel approximately is 0.05 cm^2/s, and thickness values for many coupons are in the 0.1–0.3 cm range [32]. For an axial Fourier number of 0.3, the corresponding times

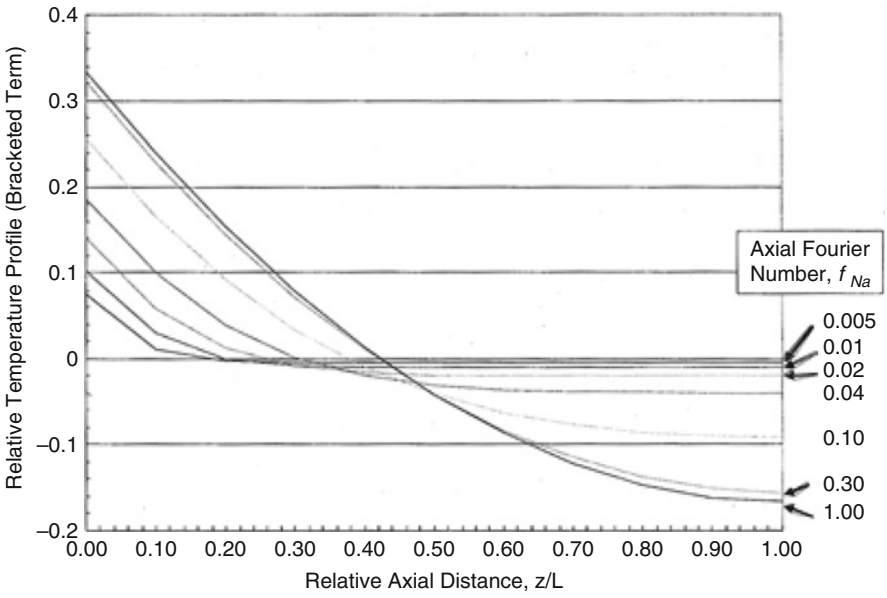


Fig. 6.20 Relative axial temperature distribution in a heated finite plate as a function of axial Fourier number

are in the 0.06–0.54 s range. This means that, for many cases of interest (exposure time greater than 0.5 s at these thicknesses), Equation 6.19 may be reduced to a simpler form for analyzing plate heating,

$$T(z,t) = \frac{AIt}{\rho C_{\mathrm{p}} L} + \frac{AIL}{k}\left\{\frac{3(1-z/L)^2 - 1}{6}\right\} \tag{6.20}$$

Evaluating this equation at $z = 0$ and $z = L$, the following two results are obtained:

$$T(0,t) = T_0(t) = \frac{AI}{\rho C_{\mathrm{p}} L}\left[t + \frac{L^2}{3\kappa}\right] + T_0 \tag{6.21}$$

$$T(L,t) = T_L(t) = \frac{AI}{\rho C_{\mathrm{p}} L}\left[t + \frac{L^2}{6\kappa}\right] + T_0 \tag{6.22}$$

where advantage has been taken of the relationship between thermal conductivity and diffusivity: $k = \rho C_{\mathrm{p}} \kappa$. Taking the first derivative with respect to time of Eq. 6.22 and rearranging, we arrive at:

$$A = \frac{\rho C_{\mathrm{p}} L}{I}\frac{\mathrm{d}T_L}{\mathrm{d}t} \tag{6.23}$$

Thus, given knowledge of the density, specific heat, and local irradiance (opposite the location where the temperature is being measured), the absorptance can be determined from the slope of the rear surface's thermal response curve.

For most materials, the specific heat is temperature dependent. In addition, the irradiance may vary with time. Then Eq. 6.23 can be conveniently evaluated using a spreadsheet that implements measured values for $C_{\mathrm{p}}(T)$ and $I(t)$. Often the former are given by stepwise polynomial fits.

Note that this method of measuring absorptance is valid only up to the point where the target's front surface reaches its melting point. Also, the method must be applied under conditions where losses (convective, radiative, and lateral conductive) are small compared to the absorbed flux. Finally, the method should be applied when the beam's spatial irradiance profile is smooth with no local excursions about the mean greater than about 10 % [32].

6.7.4 Mathematical Modeling of Thermal Approach

The heat flow in a finite or semi-infinite thin slab for one-dimensional is governed by the partial differential equation (PDE):

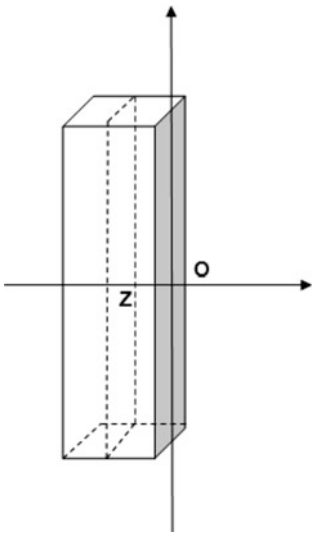


Fig. 6.21 Finite slab geometry

$$\frac{\partial u(z,t)}{\partial t} = c\frac{\partial^2 u(z,t)}{\partial z^2} \tag{6.24}$$

where c is a constant (called the diffusivity), and $u(z,t)$ is the temperature at position z and time t. The temperature over a cross section at z is taken to be uniform (see Fig. 6.21).

Many different scenarios can arise in the solution of the heat equation; we will consider several to illustrate the various techniques involved:

Example 1. Solve the following heat transfer problem along with boundary condition (BC) and initial condition (IC) as follows (for simplicity of this analysis, we have assumed $c = 1$):

$\frac{\partial^2 u(z,t)}{\partial z^2} = \frac{\partial u(z,t)}{\partial t}$

$0 < z < L$ and $t > 0$

$$\{(\text{i}) \quad u(z,0) = u_0 \quad \text{Initial Condition}$$

$$\begin{cases} (\text{ii}) \ \dfrac{\partial u(0,t)}{\partial z} = 0 \\ (\text{iii}) \ u(L,t) = u_1 \end{cases} \quad \text{Boundary Condition } \big(\text{i.e. now flow of heat over } z = 0$$

Solution Taking the Laplace transform (See Appendix E and Eq. E.38) over t of heat equation flow of above PDE gives;

$$\frac{d^2U}{dx^2} = sU - u_0$$

Then

$$U(z,s) = c_1 \cosh\sqrt{s}z + c_2 \sinh\sqrt{s}z + \frac{u_0}{s}$$

And by BC (ii), $c_2 = 0$, so that

$$U(z,s) = c_1 \cosh\sqrt{s}z + \frac{u_0}{s}$$

We find by BC (iii) that

$$U(L,s) = \frac{u_1}{s} = c_1 \cosh\sqrt{s}L + \frac{u_0}{s}$$

and so

$$c_1 = \frac{u_1 - u_0}{s\cosh\sqrt{s}L}$$

Therefore,

$$U(z,s) = \frac{(u_1 - u_0)\cosh\sqrt{s}z}{s\cosh\sqrt{s}L} + \frac{u_0}{s}$$

Taking the inverse Laplace transform gives

$$\begin{aligned} u(z,t) &= u_0 + (u_1 - u_0)\pounds^{-1}\left(\frac{\cosh\sqrt{s}z}{s\cosh\sqrt{s}L}\right) \\ &= u_1 + \frac{4(u_1 - u_0)}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n-1)}\exp\left(-(2n-1)^2\pi^2 t/4L^2\right)\cos\left(\frac{2n-1}{2L}\right)\pi z \end{aligned}$$

This is analogous to Eq. 6.19 except we have to replace $u(z,t)$ with $T(z,t)$, u_0 with $T(z,0)$ and u_1 with $T(L,t)$ as well as accounting for thermal conductivity or diffusivity c of Eq. 6.24. Very similar solution is given by Carslaw and Jaeger page 112 equation (3) and (4).

References

1. Konyukhov VK, Matrasov IV, Prokhorov AM, Shalunov DT, Shirokov NN (1970) JETP Lett 12:321

2. Hecht J (1984) Beam weapons, the next arms race. Plenum Publishing Corporation, New York
3. Davies SC, Brock JR (1987) Appl Opt 26:786
4. Menzel DH (ed) (1960) Fundamental formulas of physics, vol 2. Dover, New York, p 416, p. 416
5. Mendoza M, Jone W Effects of atmospheric turbulence on laser propagation. University of Maryland, The Maryland Optics Group
6. Sprangle P, Peñano J, Hafizi B Optimum wavelength and power for efficient laser propagation in various atmospheric environments. Naval Research Laboratory, Plasma Division
7. Smith DC (1977) High-power laser propagation - thermal blooming. Proc IEEE 65:1679
8. The infrared and electro-optical systems handbook, vol. 2, edited by F.G. Smith, Environmental Research Institute of Michigan, Ann Arbor, MI, and SPIE Optical Engineering Press, Bellingham, WA (1993)
9. Measures RM (1992) Laser remote sensing, fundamentals and applications. Krieger Publishing, Malabar, FL
10. Williams FA (1965) Int J Heat Mass Transfer 8:575
11. Caledonia GE, Teare JD (1977) J Heat Transfer 99:281
12. Armstrong RL (1984) Appl Opt 23:148
13. Armstrong RL (1984) J Appl Phys 56:2142
14. Armstrong RL, Gerstl SAW, Zardecki A (1985) J Opt Soc Am A 2:1739
15. Hänel G (1971) Beiträge zur Physik der Atmosphäre 44:137
16. Reid JS, Westphal DL, Paulus RM, Tsay S, van Eijk A (2004) Preliminary evaluation of the impacts of aerosol particles on laser performance in the coastal marine boundary layer. Naval Research Laboratory, Monterey, CA 93943-5502, NRL/MR/7534—04-8803
17. Gebhardt FG, Smith DC (1971) IEEE J Quantum Electron QE-7:63
18. Brwon RT, Smith DC (1975) Aerosol-induced thermal blooming. J Appl Phys 46(1):402
19. Hodges JA (1972) Appl Opt 11:2304
20. Johnson CS, Gabriel DA Jr (1981) Laser light scattering. Dover Publication, New York, Copyright 1981 by CRC Press, Inc
21. Weichel H (1990) Laser beam propagation in the atmosphere, Tutorial texts in optical engineering. SPIE, The International Society for Optical Engineering, Bellingham, WA
22. Ross D (1969) Light amplification and oscillators. Academic, New York, p 72
23. Bonch Bruevich M, Imas YaA (1967) Zh Tekh Fiz 37:1917(English transl.: Sov Phys-Tech Phys 12:1407 (1968))
24. Greenwood D, Primmerman C (1992) Adaptive optics research at Lincoln laboratory. Lincoln Lab J 5(1):131–150
25. Karr T (1989) Thermal blooming compensation instabilities. J Opt Soc Am A 6(7):1038–1048
26. Wall III JE (1994) Adaptive optics for high energy laser systems. Partial fulfillment of the requirements for the Degree of Master of Science in Electrical Engineering and Computer Science, Massachusetts Institute of Technology, May 30, 1994
27. Lukin VP, Fortes BV (2002) Adaptive beaming and imaging in the turbulent atmosphere. SPIE PRESS, Bellingham, WA
28. Vorob'iev VV (1978) Thermal blooming of laser beams in the atmosphere: theory and model experiment. Nauka, Moscow
29. Pearson JE, Freeman RH, Reynolds HC Jr (1979) Adaptive optics techniques for wavefront correction, vol VII, Applied optics and optical engineering. Academic, New York, pp 246–340
30. Freeman RH, Pearson JE (1982) Deformable mirror for all seasons and reasons. Appl Opt 21 (4):580–588
31. http://sharp.bu.edu/~slehar/PhaseConjugate/PhaseConjugate.html
32. Accetta JS, Loomis DN (2008) High energy laser (HEL) lethality data collection standards-revision A. Directed Energy Professional Society, Albuquerque, New Mexico

Chapter 7
Lasers

The word laser is an acronym for light amplification by stimulated emission of radiation, although common usage today is to use the word as a noun—laser—rather than as an acronym—LASER.

A laser is a device that creates and amplifies a narrow, intense beam of coherent light.

Atoms emit radiation. We see it every day when the "excited" neon atoms in a neon sign emit light. Normally, they radiate their light in random directions at random times. The result is incoherent light—a technical term for what you would consider a jumble of photons going in all directions.

The trick in generating coherent light—of a single or just a few frequencies going in one precise direction—is to find the right atoms with the right internal storage mechanisms and create an environment in which they can all cooperate—to give out their light at the right time and all in the same direction.

7.1 Introduction

In this chapter we discuss the response of materials to a high-power radiation laser with a one-dimensional mathematical modeling and presentation of solution to different cases of heat conduction partial differential equation along with given boundary and initial conditions under different scenarios. The topics of this chapter also include discussion of optical reflectivity of metals at Infrared Radiation (IR) wavelengths, laser-induced heat flow in materials, the effects of melting and vaporization, the impulse generated in materials by both pulsed and continuous radiation, and the influence of the absorption of laser radiation in the blow-off and melting region in front of the irradiated materials.

LASER is an acronym that standing for "Light Amplification by the Stimulated Emission of Radiation," which is almost totally a coherent beam. In order to produce a lasing beam we look at an atom that will emit a photon of light when it

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_7

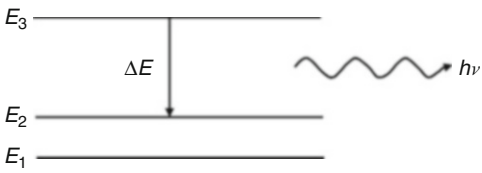


Fig. 7.1 Energy levels

decays from an excited energy state to a lower state. The energy difference between two states can be designated as ΔE which can be described in terms of frequency ν of this light in the form of Eq. 7.1, which is also known as Planck–Einstein Relation in quantum mechanics, for the above photon emitting light due to its decay in energy level.

$$\Delta E = h\nu \tag{7.1}$$

where h is Planck's constant and its value is given as $h = 6.62606957 \times 10^{-34}\text{J s} = 4.135667516 \times 10^{-15}\text{eV s}$. Equation 7.1 in terms of energy level is depicted below (Fig. 7.1).

If the wavelength of emitted photon light is designated with symbol λ and speed of light with c, then the relationship between wavelength and speed of light can be written as:

$$c = \lambda\nu \tag{7.2}$$

Substitution of Eq. 7.2 into Eq. 7.1 for frequency ν, we get a new expression for the Planck–Einstein relationship as:

$$\Delta E = \frac{hc}{\lambda} \tag{7.3}$$

The above equation leads to another relationship involving Planck's constant h. Given p for the linear momentum of a particle (not only a photon, but other particles as well), the de Broglie wavelength λ of the particle is given by:

$$\lambda = \frac{h}{p} \tag{7.4}$$

In some applications where it makes sense to use the angular frequency, where the frequency is expressed in terms of radians per second instead of rotation per second or Hertz, it is customary to absorb a factor of 2π into the Planck constant. The resulting constant is called *Reduced Planck Constant* or *Dirac Constant*. It is equal to the Planck constant divided by 2π, and is denoted as $\hbar$ and pronounced ***h-bar***.

$$\hbar = \frac{h}{2\pi} \tag{7.5}$$

Therefore, the energy of photon with angular frequency ω, where $\omega = 2\pi v$, is given as

$$\Delta E = \hbar\omega \tag{7.6}$$

The reduced Planck's constant is the quantum of angular momentum in quantum mechanics.

The numerical value of reduced Planck constant is given as $\hbar = 1.054571726 \times 10^{-34}$ J s $= 6.58211928 \times 10^{-16}$eV s.

The above conditions and circumstances is the case for any light source, whether laser, flame, incandescent body, etc.

Atoms emit photons for any conventional light source in a random mode, sporadic manner and spontaneously decay to a lower energy state when are excited by heat or any other heat-generated source, such as electric current. On the other hand, physics of laser indicates that the photons are emitted in phase and the electromagnetic radiation behavior types are encountered and more or less, we can describe it as a simply wave propagation of sinusoidal radiation filed takes place and at a microscopic level can be defined by the following mathematical solution of wave equation in conductor and taking the real part of the solution of the wave equation (i.e., we assumed the general solution is complex quantity type that includes both real and imaginary terms as part of solution) under consideration, then the relationship is presented as below:

$$\varepsilon(z,t) = \Re e\left[\varepsilon_0 \mathrm{e}^{-2\pi kz/\lambda} \mathrm{e}^{i\omega(t-nz/c)}\right] \tag{7.7}$$

where

$\varepsilon =$ is the electric field of the radiation
$\Re e =$ stands for the real part of the complex quantity in brackets
$\varepsilon_0 =$ is the maximum amplitude
$k =$ is the extinction coefficient and vacuum, $k = 0$
$z =$ is the direction in which the wave is propagating
$\lambda =$ is the wavelength
$t =$ is time
$n =$ is the index of refraction and in vacuum $n = 1$
$c =$ is the speed of light in vacuum

Equation 7.7 is just Electric Field solution to set of Maxwell's Equation inside a linear, homogeneous and isotropic conducting medium that has electric permittivity ε and magnetic permeability μ. The solution in general form of vector presentation for electric field using complex notation is:

$$\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 \mathrm{e}^{(-kz)} \mathrm{e}^{[i(kz-\omega t)]} \tag{7.8}$$

Readers can refer to Appendix F to find how the solution of wave equation in conductor will result in Eq. 7.7 in above, which is nothing more than the standard representation of the electric field of traveling light wave. You can also look at the quick approach to derive the Eq. 7.7 in next page under Wave Equation as well

Wave Equation
The macroscopic Maxwell's Equations read as follow:

$$\vec{\nabla} \times \vec{E} = \frac{\partial \vec{B}}{\partial t} \tag{1}$$

$$\vec{\nabla} \times \vec{H} = \vec{j} + \frac{\partial \vec{D}}{\partial t} \tag{2}$$

$$\vec{\nabla} \cdot \vec{D} = \rho \tag{3}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{4}$$

The relationship between $\vec{E}$ (Electric Field), $\vec{D}$ (Electric Displacement), $\vec{B}$ (Magnetic Field), and $\vec{H}$ (Magnetic Field Strength) generally speaking is nonlinear, but in our case of interest for high-power laser interaction with materials we can approximate them by a linear models and the relationship in general will depend on the frequency of the radiation field. Note that these parameters and relationship between them describe the material behavior. In case of time harmonic fields the Fourier-transformed field quantities are related according to the following sets of equations:

$$\vec{D}(\vec{r}, \omega) = \varepsilon_0 \varepsilon(\omega) \vec{E}(\vec{r}, \omega) \tag{5}$$

$$\vec{B}(\vec{r}, \omega) = \mu_0 \mu(\omega) \vec{H}(\vec{r}, \omega) \tag{6}$$

With Eq. (1), (2), (5) and (6), the wave equation can be established as:

$$\vec{\nabla}\left(\vec{\nabla} \cdot \vec{E}\right) - \Delta \vec{E} = -\mu_0 \varepsilon_0 \varepsilon \frac{\partial^2 \vec{E}}{\partial t^2} \tag{7}$$

In homogeneous media and with zero space charge $\vec{\nabla} \cdot \vec{E} = 0$, and with $\mu_0 \varepsilon_0 = 1/c^2$, Eq. (7) reduces to the following form:

$$\Delta \vec{E} = \frac{\varepsilon}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} \tag{8}$$

A solution of this Partial Differential Equation [Eq. (8)] is the equation of plane wave as:

(continued)

$$\vec{E} = \vec{E}_0 e^{[i(kz-\omega t)]} \tag{9}$$

With

k—Complex wave number
ω—Real angular frequency

The complex wave number is:

$$k = k_0\sqrt{\varepsilon} = k_0 n = k_{\text{Real}} + ik_{\text{Imaginary}} = k_{\text{r}} + ik_{\text{i}} \tag{10}$$

where n is the complex index of refraction. The plane wave solution can also be cast into the form of:

$$\vec{E} = \vec{E}_0 e^{(-k_{\text{i}}z)} e^{[i(k_{\text{r}}z-\omega t)]} \tag{11}$$

If the imaginary part of the complex wave number $k_{\text{i}} > 0$ the wave decays exponentially within the material.

The Eq. 7.7 is valid relationship to measure the electric field ε in a point in space for a laser emission, where the individual photons are in phase. This equation also allows not only to measure the electric filed ε at a point in space for laser light, it will also results in predicating the oscillating of ε as well. This is not true for an light from a conventional source, and the Eq. 7.7 does not hold for the measuring such electric field at some point in space in order to express the sinusoidal variation, for the atoms emitting the light that are doing so in random, and the sinusoidal variation due to the emission from each atom needs to be averaged out to some, time-dependent value.

The fact that laser is a very coherent source of light where this coherency is created by taking advantage of simulated emission in materials in which metastable states can be induced, then by selected rules of quantum mechanics, we know that the lifetime of an atom in an excited energy state depends on these rules for transition to a lower state. Bear in your mind that, there are state from which transition to a lower level is extremely impossible and such states are called metastable states and an atom that is not going to be distributed by outside influence, will remain in a metastable state for a very long time.

If a metastable atom interacts with a photon of frequency such that Eq. 7.1 holds, where as we said ΔE is the energy difference between the atom's normal and metastable states, stimulated emission will occur. The atom will decay to its normal state by emitting another photon of frequency v, so that the net result is two photons, and the second photon will have the same phase temporally and spatially as the first.

7.2 How Laser Works

LASER is an acronym for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation. Spontaneous emission is the process by which an excited atom spontaneously emits a photon. Electrons go from excited to a resting state when a photon of energy is released. Photon emission can be stimulated by an external source of energy that will increase the population of excited electrons, a process known as pumping. A laser contains a laser chamber, a lasing medium (solid, liquid, or gas) and an external source of energy. Stimulated emission occurs when the external source of energy causes electrons to be excited in the lasing medium. A cascade reaction is generated when these excited electrons release photons, which then collide with other excited electrons in the lasing medium and cause a release of many identical photons at the same time. Laser light continues to be generated as long as the above cascade perpetuates.

Laser light has the following properties:

1. **Coherence**: Laser beam are both temporally and spatially coherent. This phenomenon results from stimulated emission, and allows laser beams to have a high-power density
2. **Collimation**: Laser beams are parallel to each other, (i.e., ignoring thermal blooming while traveling through atmospheric environment for a high energy beam), and therefore exhibit collimation. A collimated beam is created in the laser chamber when light is reflected between two mirrors and only the exit of parallel waves is allowed. Collimation allows laser light to travel long distance without loss of intensity. In practice, a lens on a laser focuses the parallel light beam down to smallest possible spot size, or the diffraction-limited spot, to allow the light to focus on the target.
3. **Radiometry**: The four main concepts in understanding laser light and target interactions are
   (a) Energy,
   (b) Power,
   (c) Fluence, and
   (d) Irradiance

   The amount of light emitted from a laser can be quantified by both energy and power. Energy represents work that is measured in Joules, while power is measure in Watts or Joules per Second which is rate at which energy is expended. The intensity of the laser beam on the target is a function of the area of the target over which it is spread that is known as *Spot Size*.

   Spot Size = Cross-Sectional Area of the Laser Beam

   Fluence which is measured in Joules per Square Centimeter is the energy density of a laser beam.

$$\textbf{Fluence} = \textbf{W} \times \textbf{S}/\textbf{Cm}^2 = \textbf{J}/\textbf{Cm}^2 + \textbf{Laser output} \times \textbf{Pulse Duration}/\textbf{Spot Size}$$

Irradiance measured in Watts per Square Centimeter which refers to the power density of a continuous wave laser beam, and it is inversely proportional to the Square Root of the Radius of the Spot Size.

$$\textbf{Irradiance} = \textbf{W}/\textbf{Cm}^2 = \textbf{Laser Output}/\textbf{Spot Size}$$

Exposure time, fluence, and irradiance of a laser can be altered depending on the particular desired laser dueling target and conditions and circumstances that laser engages the target.

Laser interacts with target in four possible ways:

1. **Reflection** *R*: Takes place when light "bounces off" the target surface without entry into the target thickness secondary to difference in the refractive index at engagement point and the environment that incoming laser beam traveling through. Increasing the angle incidence increases the amount of light reflected. Damages to target surface or target itself, occurs with particular lasers if adequate reflection of laser beams occurs and there is proper protection employed.
2. **Absorption** *A*: The absorptivity is the ratio of power that is deposited within the workpiece and the power of the incident radiation
3. **Transmission** *T*: Occurs when the laser beam passes through transparent target without altering either the target surface or the light itself
4. **Scattering** *S*: This refers to fragmentation of light after it has entered the target skin, and it results from the interaction of light with varied elements that makeup target layers. When scattering occurs, light is dispersed over a larger area within the target, and the depth of penetration (Skin Depth) of the light beam is reduced at the same time.

As we said in above, if a metastable atom interacts with a photon of frequency such that $\Delta E = hv$, where ΔE is the energy difference between the atom's normal and metastable states, stimulated emission will occur. The atom will decay to its normal state by emitting another photon of frequency v, so that the net result is two photons, and the second photon will have the same phase temporally and spatially as the first.

In laser, then, one establishes a large number of atoms in metastable states and arranges the optics to increase the likelihood of stimulated emission. Schematically, a typical laser oscillator looks like Fig. 7.2. The pumping radiation (for example, light from a flash lamp) excites the atoms in the lasing medium (for example, Cr^{+++} ions in ruby).

In the decay process (if we have a successful laser), a large number of ions are left in a metastable state; this is called a population inversion. As some atoms begin to decay, they stimulate others to decay. But this alone would not provide a laser,

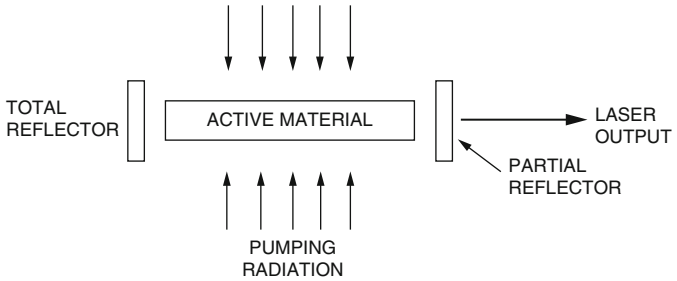


Fig. 7.2 Schematic representation of a laser

since the emission would occur in random directions. The role of reflection is very important; the photons moving perpendicular to the reflectors pass through the medium many times and on each pass more and more atoms are caused to emit. This results in the build-up of a very strong coherent light signal that travels in a single direction. Useful light output is obtained by making one of the mirrors a partial reflector [1].

It is interesting to look at a few examples of the intensity of laser light. In a typical ruby laser, the concentration [5] of Cr^{+++} ions is about 2×10^{19} cm^{-3}, and population inversions are of the order of $3 \times 10^{16}\,\text{cm}^{-3}$. Crudely speaking, we can think of creating $3 \times 10^{16}\,\text{quanta/cm}^3$ in the lasing medium. Since we have arranged the laser so that the output is in a single direction, and since photons move with the speed of light, we obtain $3 \times 10^{16} \times 3 \times 10^{10} = 9 \times 10^{26}\,\text{quanta/cm}^2\,\text{s}$ from the laser. For ruby, the lasing wavelength is 6943 Å, and since the energy of each quanta is hv, one can readily calculate that the output is about $2.5 \times 10^8\,\text{W/cm}^2$.

Before we go forward to the next step, let us pause and briefly present the Planck's Energy Density Distribution using Rayleigh's Energy Distribution by focusing on the understanding the nature of the electromagnetic radiation inside the cavity, by considering the radiation to consist of standing waves having a temperature T with nodes at the metallic surface. By arguing that these standing waves are equivalent to harmonic oscillators, for they result from the harmonic oscillations of a large number of electrical charges, electrons that are present in the walls of metallic surface of cavity. When cavity is in thermal equilibrium, the electromagnetic energy density inside the cavity is equal to the energy density of the charged particles in the walls of the cavity, and the average total energy of the radiation leaving cavity for the average energy of the oscillators along with the number of standing waves or mode of radiation in the frequency interval v to $v + dv$ is written as:

$$N(v) = \frac{8\pi v^2}{c^3} \tag{7.9}$$

where $c = 3 \times 10^8\,\mathrm{m/s}$ is the speed of light and the quantity $(8\pi v^2/c^3)\mathrm{d}v$ gives the number of modes of oscillation per unit volume in the frequency range v to $v + \mathrm{d}v$ is given by:

$$u(v,T) = N(v)\langle E\rangle = \frac{8\pi v^2}{c^3}\langle E\rangle \tag{7.10}$$

Where $\langle E\rangle$ is the average energy of the oscillators present on the walls of the cavity or of the electromagnetic radiation in that frequency interval and the temperature dependence of $u(v,T)$ are buried in $\langle E\rangle$.

Now question is how we can calculate the average energy $\langle E\rangle$?. According to classical thermodynamics and equipartition theorem, all oscillators in the cavity have the same mean energy, irrespective of their frequencies.[1]

$$\langle E\rangle = \frac{\int_0^\infty E\mathrm{e}^{-E/(kT)}dE}{\int_0^\infty \mathrm{e}^{-E/(kT)}dE} = kT \tag{7.11}$$

Where $k = 1.3807 \times 10^{-23}\,\mathrm{J/K}$ is the Boltzmann constant. An insertion of Eq. 7.11 into Eq. 7.10, leads to the Rayleigh-Jeans formula:

$$u(v,T) = \frac{8\pi v^2}{c^3}kT \tag{7.12}$$

Equation 7.12 except for low frequencies is in total disagreement with experimental data: $u(v,T)$ as given by Eq. 7.12 *diverges* for high values of ν, whereas experimentally it must be finite per Fig. 7.3. Moreover, integrating Eq. 7.12 over all frequencies, the integral *diverges* which is indication of that the cavity contains an *infinite* amount of energy. Historically, this was called the *ultraviolet catastrophe*, for Eq. 7.12 diverges for *high* frequencies within the ultraviolet range.

Now, studying the Planck's Energy Density Divergence an interpolation between Wien's rule and the Rayleigh–Jeans rule—Planck succeeded in avoiding the ultraviolet catastrophe and proposed an accurate description of blackbody radiation. He considered that the energy exchange between radiation and matter must be *discrete* rather than *continuum*. His postulation indicates that the energy radiation of frequency ν emitted by oscillating charges from the walls of the cavity must come *only* in *integer multiples* of $h\nu$ as:

[1]Using a variable change $\beta = 1/(kT)$, we have $\langle E\rangle = -\frac{\partial}{\partial\beta}\mathrm{In}\left(\int_0^\infty \mathrm{e}^{-\beta E}dE\right)$ $= -\frac{\partial}{\partial\beta}\mathrm{In}(1/\beta) = 1/\beta \equiv kT$

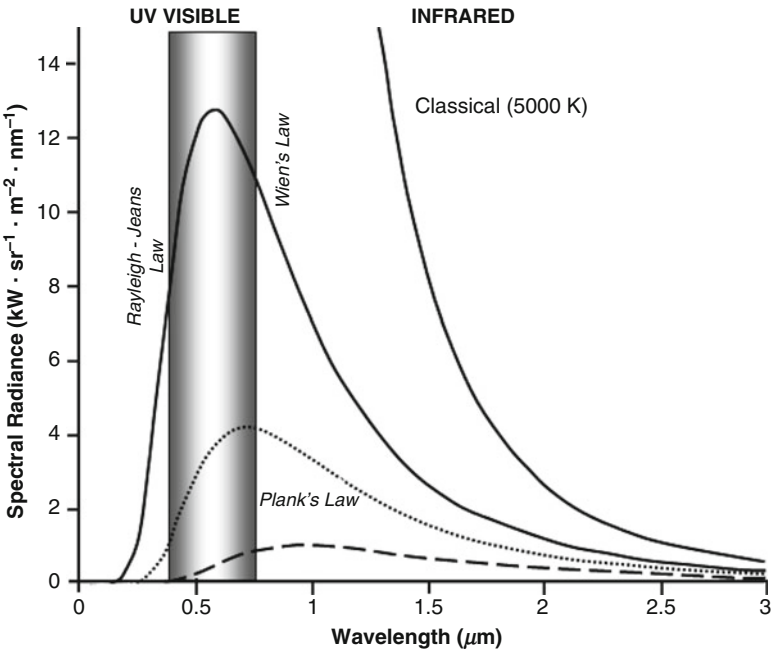


Fig. 7.3 Comparison of various spectral densities: Planck's law, *dotted curve*; Rayleigh–Jeans and Wien's law, *solid curve*

$$E = nhv \qquad n = 0,\ 1,\ 2,\ 3,\ \ldots \tag{7.13}$$

where h is a universal Planck's constant and $h\nu$ is the energy of a "quantum" of radiation, while ν represents the frequency of the oscillating charge particle in cavity's wall as well as the frequency of the radiation emitted from the walls. This is because the frequency of the radiation emitted by an oscillating charged particle is equal to the frequency of oscillation of the particle itself [8]. Equation 7.13 is known as Planck's Quantization Rule for energy or Planck's Postulate.

Therefore, assuming that the energy of an oscillator is quantized, Planck showed that the *correct* thermodynamic relation for the average energy can be obtained by merely replacing the responding to the discreteness of the oscillator's energies is given as[2]:

[2]To drive Eq. 7.14 one needs: $1/(1-x) = \sum_{n-0}^{\infty} x^n$ and $1/(1-x) = \sum_{n-0}^{\infty} nx^n$ with $x = \mathrm{e}^{-hv/kT}$

$$\langle E\rangle = \frac{\sum_{n=0}^{\infty} nh\nu e^{-nh\nu/kT}}{\sum_{n=0}^{\infty} h\nu e^{-nh\nu/kT}} = \frac{h\nu}{e^{-h\nu/kT} - 1} \tag{7.14}$$

and hence by inserting Eq. 7.14 into 7.10, the energy density per unit frequency of the radiation emitted from the hole of a cavity is given by:

$$u(\nu, T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{h\nu/kT} - 1} \tag{7.15}$$

Equation 7.15 is known as *Planck's distribution* and gives an exact fit to various experimental radiation distribution, as displayed in Fig. 7.3 and the numerical value of h obtained by fitting Eq. 7.15 with the experimental data is $h = 6.626 \times 10^{-34}$ J s.

We should note that, as shown in Fig. 7.3, we can rewrite Planck's energy density Eq. 7.15 to obtain the energy density per unit wavelength as:

$$\widetilde{u}(\lambda, T) = \frac{8\pi hc}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1} \tag{7.16}$$

In above we claimed for ruby laser of wavelength 6943 Å, the energy of each quanta is $h\nu$, one can calculate that the output is about 2.5×10^8 W/cm^2. If we compare this degree of to the power that a black body produces, let us assume sun in this case that emits at the same wavelength with a similar bandwidth, using the Planck radiation law, in terms of angular frequency $\omega = 2\pi\nu = 2\pi c/\lambda$ we can calculate:

$$U_\omega = \frac{h\omega^3}{\pi^2 c^3} \frac{1}{e^{h\omega/kT} - 1} \tag{7.17}$$

U_ω is the energy, per unit volume and per unit bandwidth, radiated by blackbody at temperature T and k is Boltzmann's constant. The radiation leaves the black-body source at rate c, so the power radiated per unit area of the source, per unit bandwidth, is

$$I_\omega = \frac{cU_\omega}{4} = \frac{h\omega^3}{\pi^2 c^2} \frac{1/4}{e^{h\omega/kt} - 1} \tag{7.18}$$

If we use the sun's temperature of 6000 K, and $\lambda = 6943$ Å, then we can show that:

$$I_\omega \approx 2 \times 10^{-5} \text{erg/cm}^2$$

For the ruby laser, a typical line width is 3 Å, so $\Delta\omega \approx 2 \times 10^{12}\ \mathrm{s}^{-1}$. Thus, the power density at the source is

$$I \approx 2 \times 10^{7} \mathrm{erg/cm}^2 s \approx 2.5\ \mathrm{W/cm}^2$$

Thus, the power density for comparable narrow-bandwidth, nearly single-frequency light is much greater at a laser source than at a conventional hot-body source, because laser light is coherent.

7.3 Laser Light Propagation

The preparation of laser light through the atmosphere posses a complex problem and it will not be discussed here. Suffice it to say that, as anyone who has driven on a foggy night certainly realizes, light is certainly scattered in the atmosphere. Lasers of high-power density poses even more difficult propagation problems because the high intensity warms the air and creates a density change across the beam. This variation in density refracts the light and causes beam spreading, or "thermal blooming."

Consider briefly the propagation of laser light in free space or in vacuum. Under these ideal conditions, the only change in the power density is due to simple beam divergence. Since the typical laser emits light that is nearly unidirectional, the beam divergence is small. In fact, one feature of a laser is that the divergence is nearly at the diffraction limit, which is of the order of λ/a, where a is the diameter of the output aperture of the laser. For the ruby laser discussed above, this gives a divergence angle of

$$\theta \approx \frac{6943 \times 10^{-8}}{1} \approx 7 \times 10^{-2}\ \mathrm{mrad}$$

for, say, a 1-cm aperture. In practice, one needs to go to much trouble to realize this limit of divergence, but it has been done. More commonly, an "off-the-shelf" ruby laser might have a beam divergence of a few mrad.

The newcomer to lasers has usually heard about diffraction-limited beams and the consequent extreme directionality of laser light lie is usually surprised to discover that at long distances front the source these beams have power densities that vary as the reciprocal of the square of the distance, like all radiating source. To see this, consider a source of power P Watt, diameter a, and divergence angle θ, as shown in Fig. 7.4.

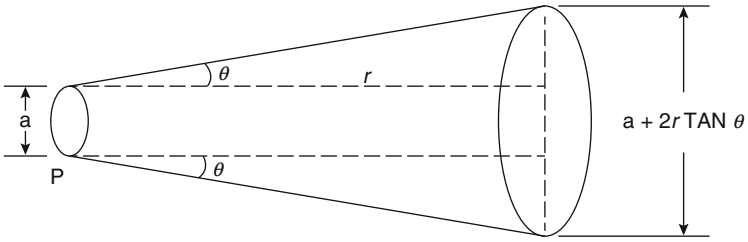


Fig. 7.4 Simplified sketch of laser beam divergence

At distance r from the source, the power density is

$$I = \frac{P}{\frac{\pi}{4}(a + 2r\tan\theta)^2}$$

or, since θ is very small and $\tan\theta \approx \theta$, then

$$I = \frac{P}{\frac{\pi}{4}(a + 2r\theta)^2}$$

or

$$I = \frac{P}{\frac{\pi a^2}{4}\left(1 + \frac{2r}{a}\theta\right)^2} \tag{7.19}$$

From this expression it is apparent that for large distance, such that $2r/a \gg 1$,

$$I = \frac{P}{\frac{\pi a^2}{4}\frac{4r^2\theta^2}{a^2}}$$

or

$$I = \frac{P}{\pi r^2\theta^2}$$

or, since $\theta \approx \lambda/a$, then we can write

$$I = \frac{P}{\pi r^2}\frac{a^2}{\lambda^2}$$

For example, consider a 10 kW beam of 10.6 μm (CO_2 laser) wavelength and 10 cm aperture at 1 mi (i.e., a high-power CO_2 laser);

$$I = \frac{10^4 \times 10^2}{\pi(5280 \times 12 \times 2.54)^2\left(10 \times 10^{-4}\right)^2}$$

or

$$I \approx 12 \ \text{W/cm}^2$$

From Eq. 7.19, if we substitute I_0, the power density at the source, for $P/(\pi a^2/4)$ and recall that $\theta \approx \lambda/a$, then we can write

$$I = I_0 \frac{1}{\left(1 + 2r\frac{\lambda}{a^2}\right)^2} \tag{7.20}$$

From this expression we can see that if r is small there is little change in power density emitted by the source. The distances at which this is true are referred to as "near field," and the fine details of the beam pattern, such as local variation in intensity, hot spots, etc., are preserved in the near field. It is apparent from Eq. 7.20 that this near-field distance will be limited to r such that $I \approx I_0$, or

$$\frac{2r\lambda}{a^2} \ll 1$$

or

$$r_{\text{near field}} \ll a^2/\lambda$$

For lasers with exceptionally good optics that have a Gaussian distribution of power density across the beam, the near field pattern will persist for distances on the order of (a^2/λ) [9–11]

As a final comment on power densities at distances from laser sources, let us use Eq. 7.20 to calculate the distance at which the power density is halved:

$$\frac{I}{I_0} = \frac{1}{2} = \frac{1}{\left(1 + \frac{2r\lambda}{a^2}\right)^2}$$

and

$$r_{1/2} = \frac{a^2}{2\lambda}\left(\sqrt{2} - 1\right)$$

For our illustration of a CO_2 laser with a 10-cm aperture, $r \approx 680$ ft, or a little more than 0.1 mi

7.4 Physics of Laser Absorption in Metals

Lasers provide the ability to accurately deliver large amounts of energy into confined regions of a material in order to achieve a desired response. For opaque materials, this energy is absorbed near the surface, modifying surface chemistry, crystal structure, and/or multi-scale morphology without altering the bulk. This chapter covers a brief introduction to the fundamental principles governing laser propagation and absorption as well as the resulting material responses.

The description of the interaction of electromagnetic radiation with matter can be done at different model level. Matter for purpose of this approach does consist of electrons and atomic kernel. Thus, the materials that laser radiation will interact with are assumed to be a linear, homogeneous and isotropic conducting medium. For spatial dimension r in this matter, we assume $r \gg 10^{-15}$ m the atomic kernels can be considered to be point charges, and for spatial dimensions greater than the classical electron radius $r \gg r_0 \sim 2.8 \times 10^{-15}$ m and the electrons can be considered to be point charges also. These point charges interact with electromagnetic fields and excite spatially and temporally fast fluctuating fields on their part. The framework of quantum electrodynamics, governs the interaction between electromagnetic fields and electrically charged particle rigorously. Therefore, the quantum theoretical treatment of electromagnetic fields leads to the concept of photons, (i.e., there can only be an integer number of photons being emitted or absorbed).

During material processing with laser radiation the result of laser treatment is predominantly determined by the power that is absorbed within the workpiece. A measure of the power that is available for the material treatment process $P_{\text{absorption}} = P_{\text{abs}}$ is the absorption A or absorptivity. The absorptivity is the ratio of power that is deposited within the workpiece and the power of the incident radiation:

$$A = \frac{P_{\text{abs}}}{P} \tag{7.21}$$

The absorptivity A can have any value between 0 and 1.0. The absorbed radiation energy is in general transformed to heat energy through heat conduction process. This energy conversion can pass several stages that can possibly be utilized for material processing. The absorption of laser radiation can for example lead to the dissociation of molecules. Before this non equilibrium state relaxes to an equilibrium state, i.e., transformation of the absorbed energy to heat energy, the dissociated molecules can be removed. Considering these conditions, then we can express the material is in ablation stage. Definition of absorptivity as such is a global value that in general contains no information that where about in the workpiece the deposition radiation energy is taking place. But we know in case of metallic interaction with laser the absorption always occurs in a thin surface layer of the metal, which is called *Localized Absorption*. Bear in your mind that this information is not included in the absorptivity A.

The absorptivity can be determined directly by measuring the power of the incident laser beam radiation and the temperature increases of the workpiece knowing the heat capacity or indirectly by measuring the power of the reflected light P_r and the power that is transmitted through the workpiece P_t as:

$$P_{abs} = P - P_r - P_t \tag{7.22}$$

Here what we mean for P_t as the transmitted power, is defined as the power that is transmitted through the whole workpiece not where P_t is the power that is transmitted through the surface of the workpiece [6].

If the radiation that is transmitted through the workpiece surface is totally absorbed within the workpiece then, the absorptivity is given by:

$$A = 1 - R \tag{7.23}$$

Where R is representation of surface reflectivity of the workpiece. In the case of non-conducting materials or very thin metal films part of the radiation that penetrates into the workpiece can leave the workpiece and we should count on transmittance part of laser beam radiation and is described by variable T, and thus, a new form of Eq. 7.23 is possible:

$$A = 1 - R - T \tag{7.24}$$

In general for heat conduction calculation and mathematical approach of high-power laser as a directed energy weapon, we can ignore the transmittance part, and therefore, Eq. 7.23 holds for purpose of these kinds of analysis. But we need to take under serious consideration of laser interaction material are parameters such as the following:

- Index of refraction n normally real part of a complex number for general index
- Index of absorption k normally imaginary part of a complex number for general index
- Electric conductivity of materials σ that laser duels with
- Heat conductivity of materials K
- Specific heat of materials at constant pressure c_p

All of the above are material-specific parameters. They solely depend on the properties of the target material at hand and the radiation wavelength can be calculated approximately for simple materials.

Furthermore, the absorptivity not only depends on the above material properties, but it is influenced by following factors as well and they are:

- Properties of the laser beam such as wavelength and polarization.
- The ambient conditions (i.e., process gas, material that surround the workpiece, etc.).
- The surface properties (i.e., roughness, morphology, etc.).

- The geometry of the workpiece (i.e., thickness, boundaries of the workpiece, etc.).
- The changes of the workpiece such as phase due to latent heat of vaporization and introduction of plasma at the ablation surface where the frequency of laser beam and plasma are equal to each other and the environment that are induced by the absorbed laser power (i.e., local heating or spot size, phase changes, laser-induced plasma).

7.4.1 Description of the Phenomena

The interaction of electromagnetic with matter can be described, by assuming that the matter consists of electrons and atomic kernels. For spatial dimensions $r \gg 10^{-15}$ m the atomic kernel can be considered as a point charge, while the electron can also be considered as a point charge because of spatial dimension greater than the classical electron radius of $r >> r_0 \tilde{} 2.8 \times 10^{-15}$. These point charges interact with electromagnetic fields or simply a propagating sinusoidal radiation field governed by Eq. 7.8 and get excited spatially and temporally fast fluctuating fields of their own part. The interaction between electromagnetic field and electrically charged particles is treated rigorously within the framework of classical quantum electrodynamics [6] as well as what is shown in Appendix F of this volume. The quantum theoretical treatment of electromagnetic fields leads to the concept of photons, i.e., there can only be an integer number of photons being emitted or absorbed (i.e., what was described in Sect. 7.2 in above). If the particles that interact with electromagnetic fields of Eq. 7.8 are bound these too can only absorbed or emit certain energy quanta. From atomistic point of view this is described as the absorption of a photon and the creation of a phonon or energy quantum of lattice oscillations.

Classical description of the spatiotemporal evolution of electromagnetic field, which is defined by the microscopic aspect of Maxwell Equations in vacuum is interacting, with these point charges of positively charged atomic kernels and negatively charged electrons.

The atoms can be defined as dipoles that are excited by the external radiation, which emits radiation on their part that interferes with the primary radiation. This process can be interpreted as coherent scattering. If radiation hits a surface of a solid matter or slab, the radiation that is emitted by the dipoles at this solid surface consists of three parts and they are as follows:

1. The first one is corresponding to the reflected wave.
2. The second one is emitted in the same direction as the incident wave and according to the Ewald-Oseen [7] extinction theorem modulus and phase of this wave are such that the incident wave and this wave extinct each other exactly within region 2. See Fig. 7.5 below.
3. The third part corresponds to the refracted wave

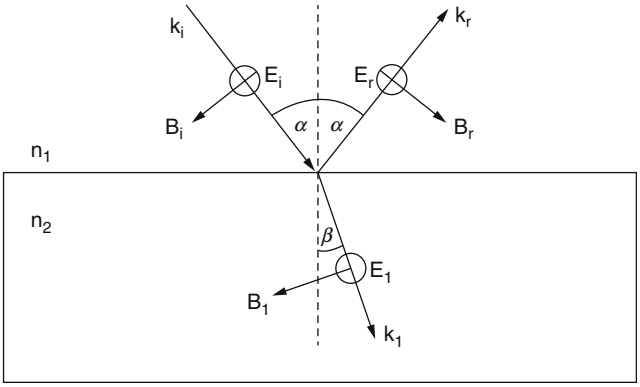


Fig. 7.5 Polarized perpendicular to the plane of incidence

Further evaluation of electromagnetic phenomena revels to average over macroscopic small but microscopic large spatial regions. Therefore, in this manner, one gets the macroscopic Maxwell's Equations. The macroscopic Maxwell equations treat the matter as a continuum whose electromagnetic properties are described by material parameters of interest target. These parameters can either be measured or calculated using microscopic approaches. The macroscopic Maxwell's Equation are presented here and solution of these sets of equation are provide in Appendix F of this volume.

$$\vec{\nabla} \times \vec{E} = \frac{\partial \vec{B}}{\partial T} \vec{\nabla} \times \vec{H} = \vec{j} + \frac{\partial \vec{D}}{\partial T} \vec{\nabla} \cdot \vec{D} = \rho \vec{\nabla} \cdot \vec{B} = 0 \tag{7.25}$$

7.5 The Behavior of Electromagnetic Radiation at Interface

One of the major advantages of the laser as a directed energy weapon for engaging target material processing is the ability to precisely control where in the material and at what rate energy is deposited. This control is exercised through the proper selection of laser processing parameters to achieve the desired material modification. In this section, we discuss the principles and equations that describe the propagation and absorption of laser energy and heat flow.

Most of the processes that occur during the interaction of laser radiation with matter start at the surfaces. For small intensities the interaction of the electromagnetic field with the target surface is driven by Fresnel formula. The Fresnel formula describes the reflection and transmission of a plane harmonic wave incident on an infinitely extended ideal plane surface. Reflection r and transmission t of the field amplitudes are defined by the following relations:

$$r = \frac{E_r}{E_i} \tag{7.26}$$

$$t = \frac{E_t}{E_i} \tag{7.27}$$

with

E_i = Electric field of the incident wave
E_r = Electric field of the reflected wave
E_t = Electric field of the transmitted wave

Although the Fresnel formula is establishing both Eqs. 7.26 and 7.27 and behavior at the surface of the metal target, but is not capable describing anything about transmission or absorption within specimen.

The Fresnel formula is driven via Maxwell's Equation sets of Eqs. 7.25 (see Appendix F as well), considering the boundary conditions for the fields at the surface of interaction with target materials. This interaction can be divided into two groups of *Perpendicular* ($\perp$) and *Parallel* ($\parallel$) polarization that is also known as s-polarized Senkrecht = Perpendicular) and p-polarized (Parallel) as well and we explain them further down in this chapter.

In case of perpendicular polarization the electric field vector is perpendicular to the plane that is spanned by the incident and reflected wave vectors, whereas in case of parallel polarization the field vectors are parallel to this plane. The wave vector $\vec{k}_i$, $\vec{k}_r$, and $\vec{k}_t$ all lie in one plane. The vector of the electric field can have components perpendicular ($\perp$) or parallel ($\parallel$) to this plane. Reflection and transmission depend on the direction of the electric field vector relative to the plane of incidence as it illustrated in Fig. 7.6 below.

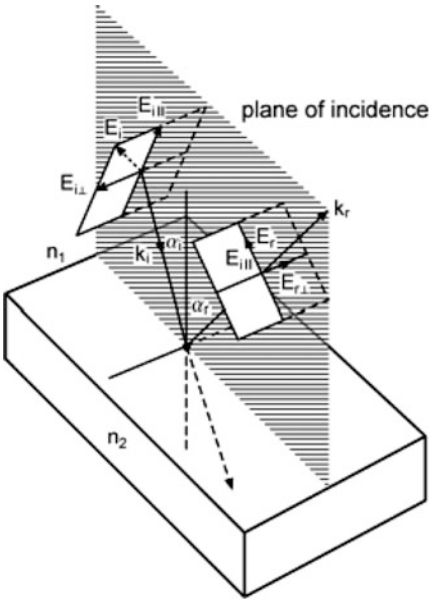


Fig. 7.6 Incident, reflected, and transmitted beams lie in one plane. The electric field vector can be split into two components: one component is parallel and the other component is perpendicular to this plane

For practical applications the polarization of the laser radiation is often chosen so that the field strength is either perpendicular or parallel to the plane of incidence and very special cases circular polarization is used. In these cases, the decomposition of the field vectors in their perpendicular and parallel components can be omitted and the mathematical treatment is simplified. Figure 7.5 shows the situation in the case of normal incidence.

The angle α that the wave vectors of the incident and reflected wave, respectively, make with the surface normal and angle β of refracted wave are connected by Snell's law as:

$$n_1 \sin\alpha = n_2 \sin\beta \tag{7.28}$$

n_1 and n_2 are the indices of refraction of the two media. The Fresnel formula for perpendicular polarization read

$$\frac{E_\mathrm{r}}{E_\mathrm{i}} = r_\mathrm{s} = \frac{n_1 \cos\alpha - \dfrac{\mu_1}{\mu_2}\sqrt{n_2^2 - n_1^2 \sin^2\alpha}}{n_1 \cos\alpha + \dfrac{\mu_1}{\mu_2}\sqrt{n_2^2 - n_1^2 \sin^2\alpha}} \tag{7.29}$$

$$\frac{E_\mathrm{t}}{E_\mathrm{i}} = t_\mathrm{s} = \frac{2n_1 \cos\alpha}{n_1 \cos\alpha + \dfrac{\mu_1}{\mu_2}\sqrt{n_2^2 - n_1^2 \sin^2\alpha}} \tag{7.30}$$

with μ_1 and μ_2 for magnetic permeability of media 1 and 2 respectively.

Figure 7.7 shows the situation for parallel polarization. The Fresnel formula for parallel polarization read (**see Appendix G, Optics for derivation**) as:

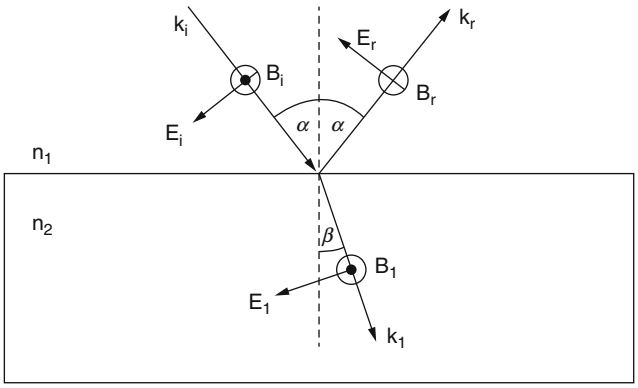


Fig. 7.7 The field vector of the reflected wave is often given in the opposite direction. In that case Eq. 7.32 must be multiplied by -1

$$\frac{E_r}{E_i} = r_p = \frac{\frac{\mu_1}{\mu_2} n_2^2 \cos\alpha - n_1 \sqrt{n_2^2 - n_1^2 \sin^2\alpha}}{\frac{\mu_1}{\mu_2} n_2^2 \cos\alpha + n_1 \sqrt{n_2^2 - n_1^2 \sin^2\alpha}} \tag{7.31}$$

$$\frac{E_t}{E_i} = t_p = \frac{2 n_1 n_2 \cos\alpha}{\frac{\mu_1}{\mu_2} n_2^2 \cos\alpha + n_1 \sqrt{n_2^2 - n_1^2 \sin^2\alpha}} \tag{7.32}$$

In the case of normal incidence, i.e., $\alpha = 0$, the plane of incidence cannot be defined uniquely any more and the difference between perpendicular and parallel polarization vanishes. In that case Eqs. 7.29 and 7.31 and Eqs. 7.30 and 7.32 give the same results. The amplitudes of the reflected and transmitted waves in case of normal incidence are given by:

$$\frac{E_r}{E_i} = \frac{n_1 - n_2}{n_2 + n_1} \tag{7.33}$$

$$\frac{E_t}{E_i} = \frac{2 n_1}{n_2 + n_1} \tag{7.34}$$

When $n_2 > n_1$ the reflected wave undergoes a phase change of 180° and this holds for real indices of refraction.

7.5.1 Light Propagation in Materials

The primary interaction between laser radiation and a solid is photochemical excitation of electrons from their equilibrium states to some excited states by the absorption of photons.

Some of these transitions are schematically shown in Fig. 7.8. Interband transitions take place when photon energy is larger than band gap of the material. In this process, electron–hole pairs are generated. The free electrons may jump back from conduction band to valence band through thermal (dashed lines) or photochemical processes. If the photon energy is less than band gap of the material, the energy can be absorbed by defect levels in the band gap or produce Intra-band transitions. Both transitions will induce thermal processes as electrons jump back to valence band. With higher laser light intensities, multi-photon absorption is favored, because the probability of nonlinear absorption increases strongly with laser intensity. The coherent multi-photon transitions would generate electron–hole pairs similar to inter-band transitions.

Thus, the initial electronic excitation is followed by complex secondary processes, which can be classified into thermal and photochemical processes. The type of interaction between laser radiation and the material depends on laser parameters (wavelength, pulse duration, and fluence) and on the properties of the materials [26].

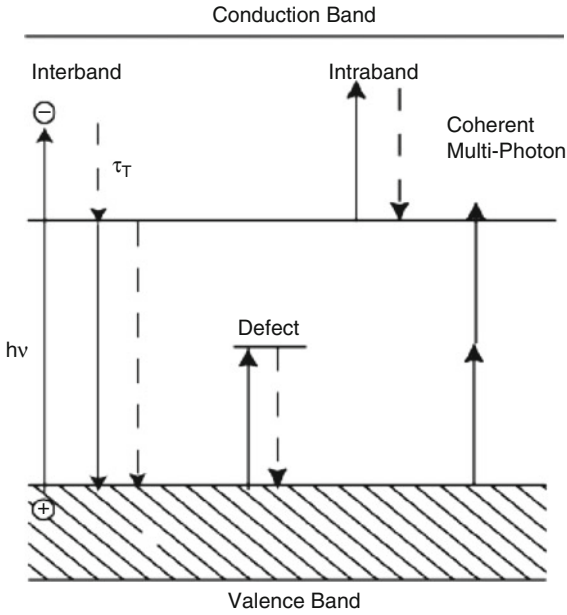


Fig. 7.8 Schematic of different types of electronic excitation in a solid

Laser ablation (material removal) can be analyzed on the basis of photothermal (purely photolytic), photochemical (purely photolytic), and photo-physical processes, in which both thermal and non-thermal mechanisms contribute to the overall ablation rate.

Confinement of deposited energy to desired regions on a material's surface can be achieved by controlling the laser's spatial intensity profile. The predominant methods for control include beam steering by fixed or galvanometric scanning mirrors, beam focusing through telescoping or converging optics, and beam shaping with homogenizers [12], amplitude masks, refractive elements [13], and diffractive optical elements [13]. However, one can also use more advanced optical devices such as spatial light modulators [14], deformable mirrors [15], and tunable acoustic gradient index (TAG) lenses [16] allowing real-time modulation of the beam's intensity profile on the surface. There has been extensive work in the area of beam shaping with a number of articles and books [17, 18], as well as a chapter in this book devoted to the subject.

When light strikes the surface of a material, a portion will be reflected from the interface due to the discontinuity in the real index of refraction and the rest will be transmitted into the material. The fraction of the incident power that is reflected from the surface R depends on the polarization and angle of incidence θ_i of the light as well as the index of refraction of the atmosphere n_1 and the material n_2. The reflection coefficients for the s-polarized and p-polarized components of the light can be calculated from the well-known Fresnel equations [19], which we provided

as Eqs. 7.29–7.34 for normal incidents as well and now we rewrite it in a different form below [19]:

$$R_s = \left[\frac{E_r}{E_i}\right]^2 = \left[\frac{n_1 \cos(\theta_i) - n_2 \cos(\theta_t)}{n_1 \cos(\theta_i) + n_2 \cos(\theta_t)}\right]^2 \tag{7.35a}$$

$$R_p = \left[\frac{E_r}{E_i}\right]^2 = \left[\frac{n_1 \cos(\theta_t) - n_2 \cos(\theta_i)}{n_1 \cos(\theta_t) + n_2 \cos(\theta_i)}\right]^2 \tag{7.36a}$$

and are related to the transmission coefficients through $T_s = 1 - R_s$ and $T_p = 1 - R_p$.

For perfectly flat surface both Eqs. 7.35a and 7.36a, produce the following forms of equation sets 7.35b and 7.36b, assuming media 1 is vacuum so $n_1 = 1$ and $n_2 = n$ for media or target index of refraction with φ being angle of incidence and parameters such as $\kappa =$ thermal diffusivity is target material property:

$$R_s = \frac{(n - \cos\varphi)^2 + \kappa^2}{(n + \cos\varphi)^2 + \kappa^2} \tag{7.35b}$$

$$R_p = \frac{\left(n - \frac{1}{\cos\varphi}\right)^2 + \kappa^2}{\left(n + \frac{1}{\cos\varphi}\right)^2 + \kappa^2} \tag{7.36b}$$

For the case of normally incident light on a flat surface, the above equations reduce to the more familiar form:

$$R = R_s = R_p = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \tag{7.37}$$

which is analogous to Eq. 7.33, in above.

A variation of the reflectivity with angle of incidence is shown in Fig. 7.9. At certain angles the surface electrons may be constrained from vibrating. Otherwise electrons would have to leave the surface and they would be unable to do that (collective vibrational modes) without disturbing the matrix, i.e., absorbing the photon. Thus, if the electric vector is in the plane of incidence, the vibration of the electron is inclined to interfere with the surface and absorption is thus high.

While if the plane is at right angles to the plane of incidence then the vibration can proceed without reference to the surface and reflection is preferred. There is particular angle—the "Brewster" angle—at which the angle of reflection is at right angles to the angle of refraction. When this occurs, it is impossible for the electric vector in the plane of incidence to be reflected since there is no component at right angles to itself. Thus, the reflected ray will have an electric vector only in the plane at right angles to the plane of incidence. At this angle the angle of refraction $= 90°$,

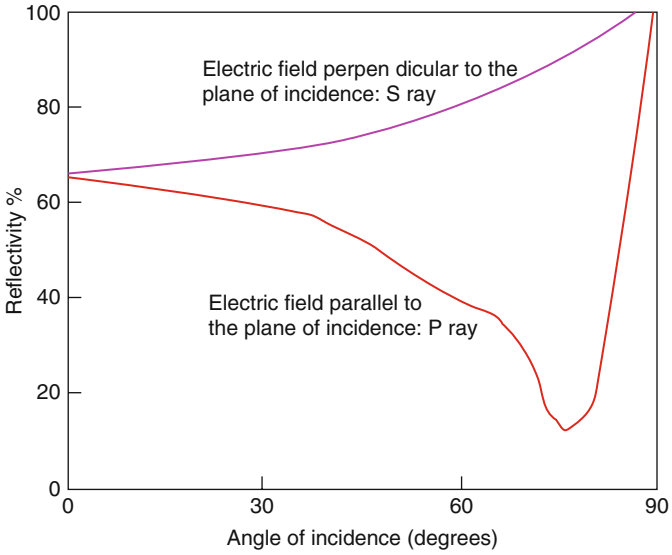


Fig. 7.9 Reflectivity of steel to polarized 1.064 μm radiations

which is the angle of incidence and hence by Snell's law the refractive index, $n = \tan$ (Brewster angle). Any beam, which has only, or principally, one plane for the electric vector is called a "polarized" beam.

The reflectivity of a given material will depend on the frequency of the light source through the dispersion relation of its index of refraction. For instance in the case of normal incidence, values for reflectivity of metals in the near UV and visible spectral range are typically between 0.4 and 0.95, and between 0.9 and 0.99 for the IR [20]. In addition, the reflectivity of a surface will depend on the temperature of the material through changes in the permittivity, band structure, plasma oscillations, or material phase [21]. For instance, upon melting, the reflectivity of silicon increases by a factor of about 2 [21], while that of a metal such as Ni changes by only a few percent [22]. In the case of small scale or structured materials, additional optical resonances are possible, such as surface and bulk plasmons and polaritons, which can lead to enhanced absorption or reflection due to the details of the photon–electron interactions [23].

Once inside the material, absorption causes the intensity of the light to decay with depth at a rate determined by the material's absorption coefficient α. In general, α is a function of wavelength and temperature, but for constant α, intensity I decays exponentially with depth z according to the Beer–Lambert law,

$$I(z) = I_0 \mathrm{e}^{-\alpha z} \tag{7.38}$$

where I_0 is the intensity just inside the surface after considering reflection loss.

The magnitude of the gradient of intensity yields the volumetric energy deposition rate $\alpha I_0 \mathrm{e}^{-\alpha z}$.

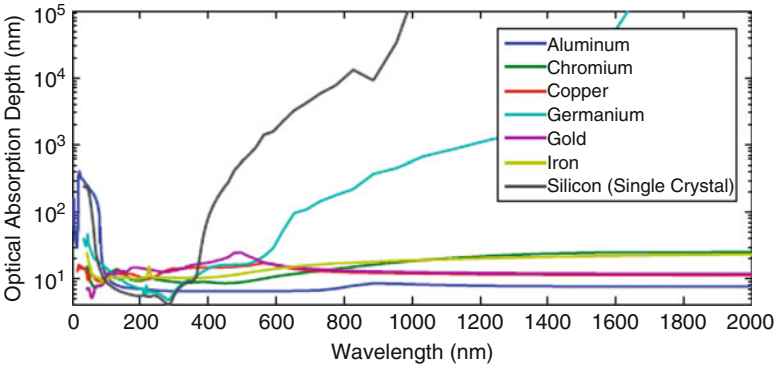


Fig. 7.10 Optical absorption depth for several materials over a range of wavelengths [27]

It is convenient to define the optical penetration or absorption depth, $\delta = 1/\alpha$ which is the depth at which the intensity of the transmitted light drops to $1/e$ of its initial value at the interface. Figure 7.10 shows optical absorption depths as a function of wavelength for a variety of metals and semiconductors. The important thing to note from Fig. 7.10 is that the absorption depths are short relative to bulk material dimensions. For instance, in the case of most metals undergoing UV illumination, the absorption depth is on the order of 10 nm. Although the interpretation of absorption depth was developed for a plane wave, the fact that energy absorption is approximately confined within the absorption depth still holds for more general beam profiles. Therefore, choosing wavelength with short absorption depths can allow local modification of surface properties without altering the bulk of the material.

The above treatments considered only linear optical phenomena; however, this is not necessarily the case in all materials, nor for all incident laser conditions. Some materials such as glasses exhibit strong nonlinearities in their index of refraction [28], which can lead to a number of interesting effects such as self-focusing, defocusing, or soliton propagation [29].When dealing with CW or nanosecond duration laser pulses, it is typically assumed that most of the absorption is due to single photon interactions. However, for picosecond (ps) and femtosecond (fs) lasers, the extremely high instantaneous intensity enables phenomena such as optical breakdown and multi-photon absorption which can significantly decrease absorption depths [29].

Another consideration of physics of high-power laser interaction with target materials is *Focal Spot Size*, which determines the maximum energy density that can be achieved when the laser beam power is set to engage target surface or workpiece and in industry this phenomena also plays an important rule for material processing.

In order to adjust the beam, to guide it to the workpiece and shape it, there are many devices. In nearly all of them the simple laws of geometric optics are

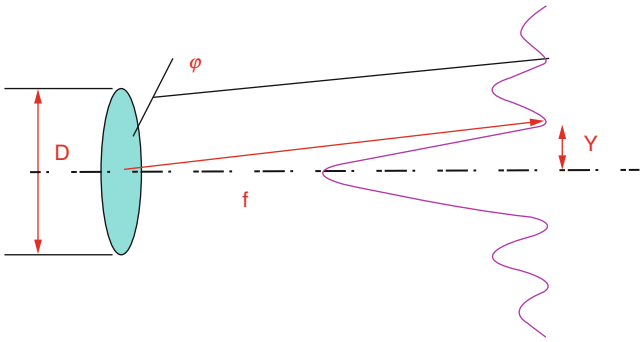


Fig. 7.11 Focus patter of parallel laser light [32]

sufficient to understand how they work. However, to calculate the precise spot size and depth of focus one needs to refer to Gaussian optics and diffraction theory.

As we discussed before (i.e., Sect. 7.3), a beam of finite diameter is focused by a lens onto a plane as shown in Fig. 7.11. When a beam of finite diameter D is focused by a lens onto a plane, the individual parts of the beam striking the lens can be imaged to be point radiators of new wavefront. The light rays passing through a lens will converge on the focal plane and interfere with each other, and thus, constructive and destructive superposition will take place and light energy is distributed as shown in Fig. 7.11. The central maximum contains about 86 % of the total power. The focusing diameter is measured between the points where the intensity has fallen to $1/e^2$ of the central peak value.

For a circular beam with a plane wavefront, the diffraction limited beam diameter, which is the smallest focal diameter ($\omega_{\min}$), that is given by Eq. 7.39:

$$\omega_{\min} = 2.44\left(\frac{f \cdot \lambda}{D_{\mathrm{L}}}\right) \tag{7.39}$$

where f is the lens focus length, λ is the light wavelength, D_{L} is the unfocused beam diameter. Equation 7.30 is representation of the upper limit of laser processing precision at a specific frequency, although in theory, laser should be free of these limitations, considering its coherency nature.

Equation 7.39, can be modified to the smallest possible focal spot size in this case as:

$$\omega_0 = 2.44\left[\frac{f \cdot \lambda}{D}\right](2p + l + 1) \tag{7.40}$$

where f is the lens focal length, D is the beam diameter at the lens, λ is wavelength of the light, p and l are the mode number.

The ideal laser should produce an intense, perfectly collimated beam of light that could be focused to a very small spot size. That tiny intense spot of light could be

used to cut any shape in any materials. Unfortunately, such a well-focused high intensity laser was hard to find before 1990, and even present ones are impractical for some laser cutting in industry.

The problem can be explained through the following Eq. 7.41 as:

$$D_{\mathrm{F}} = M^2 \frac{4}{\pi} \cdot \lambda \cdot \frac{f}{D_{\mathrm{L}}} \tag{7.41}$$

where D_{F} is the diameter of the laser beam in the focal plane of the focusing lens, M^2 is the beam quality factor, λ is wavelength of the laser light, f is the focal length of the focusing lens, and D_{L} is the diameter of the collimated laser beam on the focusing lens.

Further analysis of Eq. 7.41 indicates that if one wants smaller diameter of the laser beam D_{F}, it requires either an smaller ration of f/D_{L}, or make f very small and D_{L} very large. However, one should also consider the laser's Raleigh length R_{L},

$$R_{\mathrm{L}} = D_{\mathrm{F}} \cdot \frac{f}{D_{\mathrm{L}}} \tag{7.42}$$

which is the distance above and below the focal plane where diameter of the beam has increased by and the beam intensity has dropped by $\sqrt{2}$. In practice, one can cut very thick materials, which are as thick as Raleigh lengths. To cut thicker materials one would need large f/D_{L} ratio, so the only parameters left is the wavelength and beam quality M^2 to be adjusted to reduce the spot diameter D_{F} for the purpose of cutting very small objects with very small laser beam as a cutting tool or utilization of very high-power laser as a directed energy weapon.

Ultrafast laser oscillators generate pulse energy initially at nJ (i. e., 10^{-9} J) scale, amplification to mJ level is needed in micromachining, but one knows, when pulse lasting time is very small, peak energy intensity goes up far beyond the safe operation range of normal optical amplification systems. Ultra-short pulsed lasers successfully solved this difficulty using Chirped Pulse Amplification (CPA) and Pulse Compression techniques. The energy of ultrafast lasers are highly concentrated in time domain, but their frequency distributions are much broader than normal laser systems. Light at different frequencies travel at different speeds through optical mediums and this is not very good thing at first glance, but contrary to one's intuition, this forms the base for the final solution. In free space, different components in a broad band laser pulse travel at nearly same speed. When normal optical components are in the optical path, long wavelength light components travel through the medium faster than short wavelength components, and thus, pulse lasting time is stretched and energy intensity is lowered. Special optical devices such as chirped mirrors or special prism pairs are used to compensate the pulse spreading; they allow short light components pass through faster than long wavelength components, so they compress pulse lasting time.

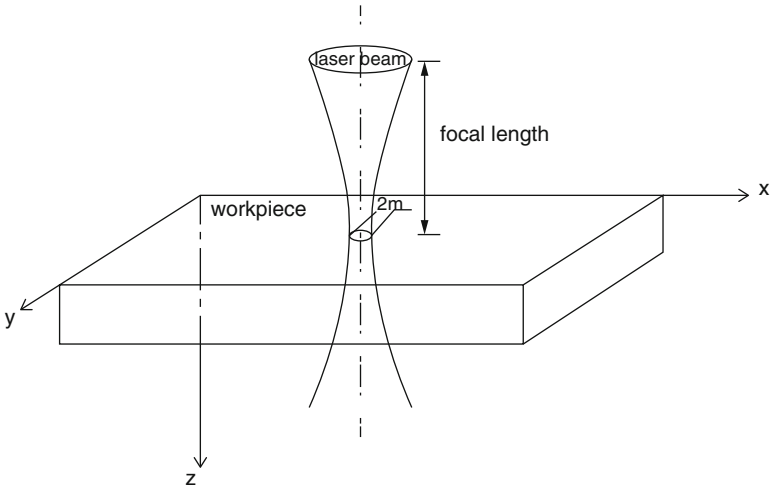


Fig. 7.12 Schematic of laser beam impinging on a workpiece

Further analysis of Eq. 7.40, clearly shows not only shows the influence of modes on the focal property, but other factors that also affect focal spot size, such as spherical aberration and thermal lens effects should be taken under consideration as well.

Most lenses are made with a spherical shape, but they cannot be of perfect shape because there exists spherical aberration. Lenses in laser systems transmit or reflect high-power laser radiation. However, laser power variations can cause shape changes of the lenses, so the focal point will change when the radiation power changes, thus affecting the focal spot size.

According to Lasag [33], manufacturer of Nd-YAG laser system, the spot diameter of laser beam focused onto the workpiece and Fig. 7.12 below, can be approximated as:

$$2\omega \approx \frac{2f\theta}{M} \tag{7.43}$$

where f is the focal length of the lens, θ is the divergence of the laser beam before the expander, and M is the expansion factor of the beam expander.

7.5.2 Depth of Focus

The laser light is first converged at the lens focal plane, and then diverged to a wider beam diameter again. The *Depth Of Focus* (DOF) is the distance over which the focused beam has about the same intensity. This is also called beam waist,

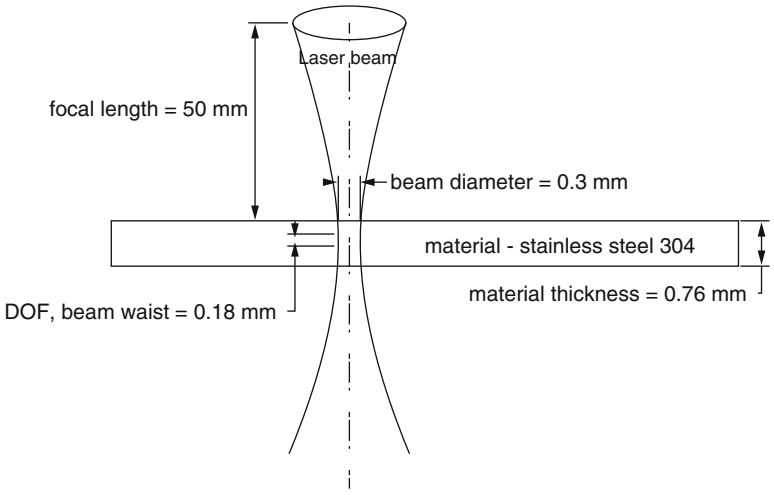

Laser beam
focal length = 50 mm
beam diameter = 0.3 mm
material - stainless steel 304
material thickness = 0.76 mm
DOF, beam waist = 0.18 mm

Fig. 7.13 Laser beam focusing onto the sample

Fig. 7.13. It is defined as the distance over which the focal spot size changes ±5 %. The equation for DOF is [33]

$$\mathrm{DOF} = \left(\frac{8\lambda}{\pi}\right)\left[\frac{f}{D}\right]^2 = 2.44\lambda\left[\frac{f}{D}\right]^2 \tag{7.44}$$

where λ is the wavelength, f is the lens focal length and D is the unfocussed beam diameter. Usually longer depth of focus is preferred, because equal energy density along the beam is preferred when using the laser to process materials.

Position of the focal plane is known to have an effect on final shape of the hole as well as degree of penetration. As the laser focusing spot moves up and down, the laser material interaction area Ar_F varies too, according to expression [34].

$$Ar_F = n\omega_0\left[1 + \left(\frac{z - z_\omega}{2\omega_0 f/D}\right)^2\right]^{\frac{1}{2}} \tag{7.45}$$

where $\omega = \omega_0$ is the laser beam radius at the beam waist equals to ω_0 which is the minimum radius of the laser beam. In our analytical and numerical calculations, Eq. 7.11 is proved to be the most accurate one to use.

While calculating the drilled-hole profiles defocusing of the laser beam effects should be taken into account. As a rule, the divergence of the laser beam would increase the radial heat flux on the hole-walls, by decreasing energy at the entrance of laser beam of the evolving hole. However, multiple reflections inside the hole will accumulate the energy at the bottom of the drilled hole, which was described before.

7.5.3 Laser Beam Quality

The concept of M^2 is important to describe actual propagation of laser beams. M^2 is a beam quality that measures the difference between the actual beam and the Gaussian beam. In order to find out the M^2 of a laser system, we need first measure the spot size along the laser optical axis.

Edge method is used if beam profile-meter is not available. Edge method uses a knife-edge to block the laser beam, a power-meter measures the power after the blocking of the knife-edge, by recording the 86 % and 14 % location of the full power. Subtracting the two values one gets the beam spot size at that point. This method measures the $1/e^2$ radius of the laser beam. Because the laser spot size is rather small, the relative measurement error for such dimension using edge method can be large. Usually a beam expander, or a collimator, is used to expand and "parallel" the beam, the spot size out of the collimator is several millimeters. For such dimension, knife-edge method can easily reduce the relative measurement error to less than 2 %. As illustrated by Fig. 7.14, six measurements at three different distances from the collimator are taken, Z_n, D_n, where $n = 1, 2, 3$. The distance from any chosen point along the optical axis and the spot size at that location are recorded.

The beam size at location Z_n, satisfies the following equation:

$$D_n^2 = D_0^2 + \left[\frac{4M^2\lambda}{\pi}\right]^2 \left[\frac{Z_n - Z_0}{D_0}\right] \quad \text{for } n = 1,\ 2,\ 3 \tag{46}$$

Where D_n is the beam size at location Z_n, D_0 is the beam waist, Z_0 is the beam waist location, λ is the wavelength, and M^2 is the beam quality parameter which is unknown. In this relation, λ is known, D_n and Z_n can be measured, M^2, D_0 and Z_0 are also unknowns.

Taking the measured data into Eq. 7.41, we get three highly nonlinear equations for three unknowns. One can solve these equations using, e.g., MathCAD. Knowing M^2, one can calculate the beam divergence, focused spot size, and depth of focus (DOF) as follows:

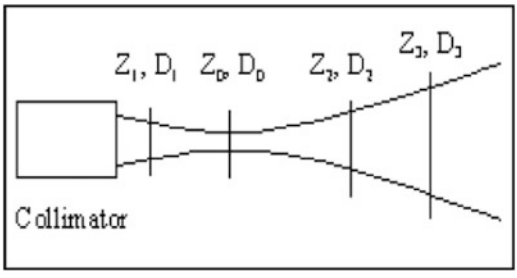


Fig. 7.14 Six different measurments at three different collimator distance

$$D_{\min} = \frac{4fM^2\lambda}{\pi D_L} \tag{7.47}$$

$$D_{\text{Gaussian}} = \left[\frac{Af}{\pi D_0}\right] \tag{7.48}$$

$$\theta_{\text{act}} = \frac{M^2\lambda}{\pi D_0}\ \theta_{\text{infinity}} = \left[\frac{D_3 - D_2}{2(Z_3 - Z_2)}\right] \tag{7.49}$$

$$\text{DOF} = \pm 0.08\lambda\left[\frac{D_{\min}}{M^2\lambda}\right] \tag{7.50}$$

where D_{L} is the laser beam size when it propagates to the front side of the focus objective lens, f is the focus length, $D_{\min}$ is the minimum beam diameter that can be achieved. θ_{act} is the real beam divergence. It can be verified that the equations are solved with an error less than 10^{-10}. In practical cases, one should first measure the pulse energy or average power of the laser beam, then measure the beam spot size along the optical axis. Using Eqs. 7.47 and 7.48 one can calculate the beam waist and beam waist location, and one can find M^2 of the beam, then can calculate the other indexes in Eq. 7.50. Knowing DOF one can calculate M^2 value as

$$M^2 = \left[0.08\frac{(D_{\min})^2}{\lambda(\text{DOF})}\right]^{1/2} \tag{7.51}$$

where $D_{\min}$ is the minimum beam diameter that can be achieved. Knowing diameter at any location, one can determine the intensity of the laser beam at that location.

7.5.4 Spherical Aberration

There are two reasons why a lens will not focus to a theoretical point. First is the diffraction limited problem as discussed in Sect. 7.3, and the second is that lens is not of a perfect shape. Most lenses are made with a spherical shaped since this can be accurately manufacture without too much cost and alignment of the beam is not so critical as with a perfect aspheric shape. The net result is that the outer ray entering the lens is brought to a shorter axial focal point than the rays nearer the center of the lens. This leaves a blur in the focal point location. The plane of best geometric focus is a little short of the plane wavefront (paraxial point). The size of the minimum spot, $d_{\min}$ is given by

$$d_{\min} = K(n,q,p)\left(\frac{D_L}{f}\right)^3 S_a = 2\Theta_a S_a \tag{7.52}$$

where Θ_a is the angular fault (half angle), S_a is distance from lens, D_L is diameter of top hat beam mode on lens, f is the focal length of the lens, $K(n,q,p)$ is the factor dependent on the refractive index n, q is the lens shape, and p is lens position,

$$K(n,q,p) = \frac{1}{128n(n-1)}\left[\frac{n+2}{n-1}q^2 + 4(n+1)pq + (3n+2)(n-1)p^2 + \frac{n^3}{n-1}\right] \tag{7.53}$$

where q is the lens shape factor $(= (r_2 + r_1)/(r_2 - r_1))$, r_1, r_2 are the radii of curvature of the two faces of the lens, and p is position factor $(= 1 - 2f/S_a)$.

7.5.5 Thermal Lens Effect

In optical elements, which transmit or reflect high-power radiation there will be some heating of the component, which will alter its refractive index and shape [35]. As the power, or the absorption, changes so will the focal point and spot size.

The two main elements usually concern are the output coupler and the focusing lens. The beam guidance mirrors could also be of concerned if adequate cooling is not supplied (water cooling or dry air cooling).

Thermal lensing is mainly caused by the rise in temperature increasing the refractive index (dn/dT) and thus shortening the focal length. A lesser effect is the thermal distortion (dl/dT). The focal length shift for thin lenses can be calculated to be

$$\Delta F = \frac{2APF^2}{\pi k D_L^2}\left(\frac{dn}{dT}\right) \tag{7.54}$$

where A is the optic absorptance, k is the thermal conductivity, T is the temperature, P is incident power of the laser beam.

Uneven heating causes further complications. The approximate Gaussian power distribution of the incident beam heats the middle more than the edge causing a radial temperature gradient (usually the edge of the lens is being cooled). A typical temperature difference would be 14° for a 1500 W beam of 38 mm diameter passing through an optic with 0.2 % absorptance.

At this time let us mention another problem. During laser micromachining, from the fact that the absorption is on the surface of the lens, there is also going to be a temperature gradient in the depth direction. The thicker the optic the more bowed will be the internal isotherms. Such an aberration will affect the M^2 value, transverse mode, and on the spatial distribution of the laser beam as well.

7.6 Theoretical Discussion of Laser Absorption and Reflectivity

This subject was touched upon in Sect. 7.4 in above and we expand it further again. To consider the coupling of the laser energy to a material, we need first to know the optical reflectivity R and the transmissivity T for light incident on a surface which divides two semi-infinite media. The transmissivity plus the reflectivity equals unity at a *single surface*:

$$R + T = 1 \tag{7.55}$$

In most practical situations we are dealing with more than one surface; typically, we have a slab of material with light impinging on one surface. Some light is reflected, and the rest is either absorbed or passed completely through the slab. In such a situation we shall describe the net result of all the reflection, after multiple passes inside the slab and appropriate absorption has been accounted for, in terms of the reflectance R, the absorptance A, and the transmittance T:

$$R + T + A = 1 \tag{7.56}$$

The absorption of the laser energy takes place through photon interaction with bound and free electrons in the material structure, which raises them to the higher energy levels. Energy conversion takes place through various collision processes involving electrons, lattice phonons, ionized impurities and defect structures.

If the surface being machined reflects too much light energy, the absorbed energy is decreased, the operation efficiency is lowered, and the reflected light may do harm to the optical systems. Reflection and absorption of laser beams is closely related to laser micromachining or dueling with target surface materials. The value of absorption and reflection is related by Eq. 7.56.

What we really are interested in from the point of view of material response is A, the absorptance of the material. In most materials of interest, from the practical aim of using lasers to melt, weld, etc., T is zero, and thus, Eq. 7.56 reduces to:

$$R + A = 1 \tag{7.57}$$

In metals, the radiation is predominantly absorbed by free electrons in an "electron gas." These free electrons are free to oscillate and reradiate without disturbing the solid atomic structure. Both Eqs. 7.56 and 7.57 are analogous to Eqs. 7.23 and 7.24 for what we found out and discussed in Sect. 7.4 in above. As an electromagnetic (*EM*) wave-front arrives at a surface of the target then all the free electrons in the surface vibrate in phase generating an electric filed 180° out of phase with the incoming beam creating "electron gas." This "electron gas" within the metal structure means that the radiation is unable to penetrate metals to any significant depth, only one to two atomic diameters (or free paths), and thus, metals are opaque and they appear shiny.

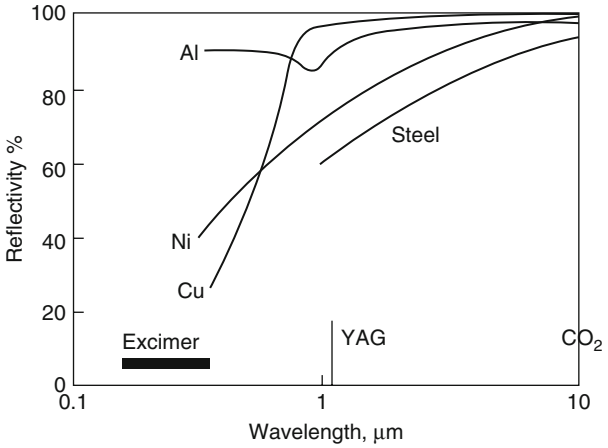


Fig. 7.15 Reflectivity as a function of wavelength for different metals

According to Fig. 7.15 reflectivity decreases as wavelength becomes shorter, while absorption increases when photon energy increases.

If sufficient energy is absorbed then the vibration becomes so intense that the molecular bonding is stretched so far that it is no longer capable of exhibiting mechanical strength and the material is said to have melted. On further heating the bonding is further loosened due to the strong molecular vibrations and the material is said to have evaporated. The vapor is still capable of absorbing the radiation, but only slightly since it will only have bound electrons. The exception occurs if the gas is sufficiently hot so that electrons are shaken free and the gas is then said to be in plasma status.

To understand reflectivity, we must use some general results from the theory of electromagnetic waves. Let us summarize these briefly at this point. The electric field of the electromagnetic wave, from Eq. 7.7, is:

$$\varepsilon(z,t) = \Re e\left[\varepsilon_0 \mathrm{e}^{-2\pi kz/\lambda}\mathrm{e}^{i\omega(t-nz/c)}\right] \quad \text{(Eq. 7.7 as before)}$$

The relationships we need are those among the index of refraction n, the extinction coefficient k, and the material properties. These relationships can be derived by substituting Eq. 7.7 in the wave equation as:

$$\frac{\partial^2 \varepsilon(z,t)}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 \varepsilon(z,t)}{\partial t^2} \tag{7.58}$$

where v is the phase velocity of the wave and for traveling wave in media of density, the phase velocity is given by

$$v = f\lambda = v\lambda \quad \text{and} \quad \omega = 2\pi f = 2\pi v = \frac{2\pi}{\lambda}$$

In case media as vacuum $v = c$ speed of light in vacuum.

If we get twice derivation of Eq. 7.7 and substitute the result in Eq. 7.58, it results in:

$$\frac{\partial^2 \varepsilon}{\partial z^2} = \mu\varepsilon \frac{\partial^2 \varepsilon}{\partial t^2} + \mu\sigma \frac{\partial \varepsilon}{\partial t} \tag{7.59a}$$

Equation 7.59a results in the expression of Eq. 7.59b below as:

$$\left(\frac{2\pi k}{\lambda} + \frac{i\omega n}{c}\right)^2 = \mu\varepsilon\left(-\omega^2\right) + i\omega\mu\sigma \tag{7.59b}$$

Note that we are using rationalized MKS units throughout. The material properties enter through μ, ε, and σ, which are the magnetic permeability, the dielectric function, and the electric conductivity of the medium. Using the usual equations between the field vectors

$$\vec{\boldsymbol{D}} = K_{\mathrm{e}}\varepsilon_0 \vec{\varepsilon} \tag{7.60a}$$

$$\vec{\boldsymbol{B}} = K_{\mathrm{m}}\mu_0 \vec{\boldsymbol{H}} \tag{7.60b}$$

$$\vec{\boldsymbol{J}} = \sigma \vec{\varepsilon} \tag{7.60c}$$

We have

$$\varepsilon = K_{\mathrm{e}}\varepsilon_0 \tag{7.60d}$$

$$\mu = K_{\mathrm{m}}\mu_0 \tag{7.60e}$$

In Eqs. 7.60a and 7.60b, ε_0 and μ_0 are the electric permittivity and magnetic permeability of a vacuum. K_{e} is the dielectric constant and K_{m} the magnetic permeability of the material. By substituting Eqs. 7.60d and 7.60e into Eq. 7.59b and using $(2\pi/\lambda) = (\omega/c)$, we obtain that:

$$(k + in)^2 = -K_{\mathrm{m}}K_{\mathrm{e}}\varepsilon_0\mu_0 c^2 + iK_{\mathrm{m}}\mu_0\sigma\left(\frac{c^2}{\omega}\right) \tag{7.61}$$

Finally, if we introduce $c^2 = (\varepsilon_0\mu_0)^{-1}$ and do some algebra, we get:

$$n - ik = \sqrt{K_{\mathrm{m}}}\sqrt{K_{\mathrm{e}} - i\frac{\sigma}{\varepsilon_0\omega}} \tag{7.62}$$

This equation relates the material parameters K_{e}, K_{m}, and σ, which in general may be complex, to index of refraction n and extinction coefficient k. To describe the

propagation of the light wave thus requires a knowledge of K_e, K_m, and σ. Before we describe these, let us look at two more general properties of our propagating electromagnetic wave.

The first of these is absorption. If the medium is absorbing, the intensity will fall off to $1/e$ of its initial value in a distance δ, obtained by setting ε^2 of Eq. 7.7 equal to $(1/e)\varepsilon_{max}^2$, or

$$\frac{4\pi k\delta}{\lambda} = 1 \tag{7.63a}$$

or

$$\delta = \frac{\lambda}{4\pi k} \tag{7.63b}$$

This shows why k is called the extinction coefficient, for it determines skin depth δ. Equation 7.63b is fairly general in that once k is known, δ can be calculated. As noted, a knowledge of the material properties is required to calculate k.

The second general property we wish to derive is the expression for reflectivity, in terms of n and k. To do this, consider light impinging normally onto an ideal solid surface, as shown in Fig. 7.64. Here we have illustrated the incident $(\vec{\varepsilon}_i)$, reflected $(\vec{\varepsilon}_r)$, and transmitted $(\vec{\varepsilon}_t)$ electric waves at a vacuum–material interface. For the present, we limit our discussion to the case of normal incidence. We now consider the boundary condition. Then, for electric field, we have

$$\varepsilon_i + \varepsilon_r = \varepsilon_t \tag{7.64}$$

For the magnetic field $\vec{\boldsymbol{B}}$, we can right:

$$B_i - B_r = B_t \tag{7.65}$$

The minus sign is before B_r because $\vec{\varepsilon} \times \vec{\boldsymbol{B}}$ is positive in the direction of propagation of the wave. Now, the relationship between $\vec{\boldsymbol{B}}$ and $\vec{\varepsilon}$, or, since $\vec{\boldsymbol{B}} = \mu\vec{\boldsymbol{H}}$, between $\vec{\boldsymbol{H}}$ and $\vec{\varepsilon}$, is required in order to proceed further (Fig. 7.16).

This follows directly from Maxwell's equations as:

$$\Delta \times \vec{\varepsilon} = -\mu\frac{\partial \vec{\boldsymbol{H}}}{\partial t} \tag{7.66}$$

$$\Delta \times \vec{\boldsymbol{H}} = \sigma\vec{\varepsilon} + \varepsilon\frac{\partial \vec{\varepsilon}}{\partial t} \tag{7.67}$$

It is convenient to rewrite Eq. 7.7 and introduce $\omega\lambda = 2\pi c$, to have $\vec{\varepsilon}$ explicitly in terms of ω instead of both ω and λ. Recall that $\vec{\varepsilon}$ is a vector, and take it as being along the x-direction. Thus, we can write:

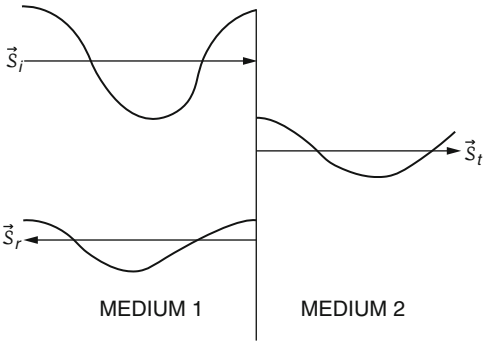


Fig. 7.16 Incident, transmitted, and reflected electric vectors at an interface

$$\vec{\varepsilon}_x = \varepsilon_0 e^{i\omega t} e^{-\frac{i\omega}{c} z(n-ik)} \tag{7.68}$$

Here we have dropped the "Re" notation, and shall simply note that we always mean the real part when we write the wave in exponential form. We shall use unit vectors $\hat{x}$, $\hat{y}$, and $\}\backslash hat\{z\}\{$.

Now the curl expression reduces to:

$$\Delta \times \vec{\varepsilon} = \hat{y} \frac{\partial E_x}{\partial z} \tag{7.69}$$

which, with Eq. 7.66, tells us that $\vec{H}$ has only a y-component, so, we have:

$$\Delta \times \vec{H} = -\hat{x} \frac{\partial H_y}{\partial t} \tag{7.70}$$

Thus, Eqs. 7.66 and 7.67 become in the following form:

$$\frac{\partial \varepsilon_x}{\partial z} = -\mu \frac{\partial H_y}{\partial t} \tag{7.71}$$

and

$$-\frac{\partial H_y}{\partial z} = \sigma \varepsilon_x + \varepsilon \frac{\partial \varepsilon_x}{\partial t} \tag{7.72}$$

and, of course, $\varepsilon_y = \varepsilon_z = H_x = H_z = 0$. Putting the expression for ε_x from Eq. 7.68 into 7.71 leads to:

$$H_y = \frac{n - ik}{\mu c} \varepsilon_0 e^{-\frac{i\omega}{c} z(n-ik)} e^{i\omega t} \tag{7.73}$$

This is desired relationship as:

$$H_y = \left(\frac{n - ik}{\mu c}\right)\varepsilon_x \tag{7.74}$$

At this point we note in passing that Eq. 7.72 or 7.59b could be used to yield the relationship of n and k to μ, ε, and σ. If the reader is unfamiliar with these relationships, it is instructive to carry out the algebra.

Returning to our consideration of the reflected electric and magnetic fields, we rewrite Eqs. 7.64 and 7.65 with the help of the relationship between H and ε, from Eq. 7.74

$$\varepsilon_\mathrm{i} + \varepsilon_\mathrm{r} = \varepsilon_\mathrm{t}$$

and

$$\mu_1 H_\mathrm{i} - \mu_1 H_\mathrm{r} = \mu_2 H_\mathrm{t}$$

become

$$\varepsilon_\mathrm{i} - \varepsilon_\mathrm{r} = \left(\frac{n_2 - ik_2}{n_1 - ik_1}\right)\varepsilon_\mathrm{t}$$

Solving for $\varepsilon_\mathrm{r}/\varepsilon_\mathrm{i}$ by eliminating ε_t, we get:

$$\frac{\varepsilon_r}{\varepsilon_i} = \frac{n_1 - n_2 - i(k_1 - k_2)}{n_1 + n_2 - i(k_1 + k_2)}$$

Finally, the reflectivity R at the surface is:

$$R = \left|\frac{\varepsilon_\mathrm{r}}{\varepsilon_\mathrm{i}}\right|^2 = \frac{(n_1 - n_2)^2 + (k_1 - k_2)^2}{(n_1 + n_2)^2 + (k_1 + k_2)^2} \tag{7.75}$$

Take medium 1 as a vacuum and drop the subscript 2, since in a vacuum $n_1 = 1$ and $k_1 = 0$, then Eq. 7.75 reduces to:

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \tag{7.76}$$

Equation 7.76 is the second relationship we will find useful in discussing the coupling of optical radiation with metals. Note that it is derived for the special case of normal incidence and is applicable to a vacuum–material interface.

7.6.1 *Reflectivity of Materials at Infrared Wavelength*

We turn now to a derivation of the optical reflectivity of metals for infrared wavelengths, where experiment has shown that the free-electron theory (sometimes called the Drude–Lorentz theory) of metals is adequate. This theory rests on three assumptions:

1. The first is that electromagnetic radiation interacts only with the free electrons in the metal.
2. The second is that the free electrons obey Ohm's law, or, more specifically, that

$$m\frac{dv}{dt} + \frac{m}{\tau}v = -e\varepsilon \tag{7.77}$$

   where m is the effective mass of the electron, v the drift velocity of electro, τ the relaxation time due to collisions (i.e., mean free time between ionic collisions. See Fig. 7.17), and $-e\varepsilon$, the force on the electron due to the electromagnetic field.
3. The third assumption is that the free electrons of a metal can be described in terms of a single effective mass, carrier concentration, and relaxation time. There has been a good deal of discussion about the validity of these assumptions in the literature.

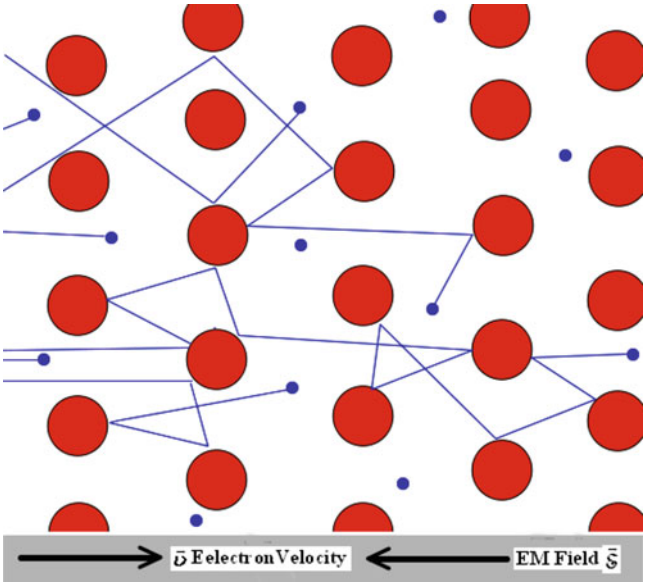


Fig. 7.17 Drude–Lorentz model electrons (shown here in *blue* or *small dot*). Constantly bounce between heavier, stationary crystal ions (shown in *red larger dot*)

Recent work [10] indicates that, for wavelengths in the intermediate infrared (a few microns to many tens of microns) and beyond, the free-electron theory does a reasonable job of predicting the reflectivity of metals.

To derive the free-electron optical reflectivity, we try solutions to Eq. 7.77 of the form

$$\upsilon \approx \mathrm{e}^{i\omega t}$$

So that Eq. 7.77 with substitution of above relation results in:

$$\left[m^*(i\omega) + \frac{m^*}{\tau}\right]\upsilon = -e\varepsilon \tag{7.78}$$

Manipulating Eq. 7.78, we can find that:

$$\upsilon = -\frac{e\tau}{m^*(1 + i\omega\tau)}\varepsilon \tag{7.79}$$

Now the current flow obeys:

$$\vec{J} = \sigma\vec{E} = -Ne\vec{\upsilon} \tag{7.80}$$

Where N is the electron concentration (number of electron per unit volume). By comparison of the last two equations,

$$\frac{\sigma}{Ne} = \frac{e\tau}{m(1 + i\omega\tau)} \tag{7.81}$$

or

$$\sigma = \frac{Ne^2\tau}{m(1 + i\omega\tau)} \tag{7.82}$$

Now the dc conductivity is

$$\sigma_0 = \frac{Ne^2\tau}{m} \tag{7.83}$$

We see σ is a complex quantity and seek to write it as the sum of a real and imaginary part. Thus,

$$\sigma = \frac{Ne^2\tau(1 - \omega\tau)}{m(1 + \omega^2\tau^2)} \tag{7.84}$$

Define

$$\sigma = \sigma_1 - i\sigma_2 \tag{7.85}$$

The result is

$$\sigma_1 = \frac{\sigma_0}{1 + \omega^2\tau^2} \tag{7.86a}$$

$$\sigma_2 = \frac{\sigma_0\omega\tau}{1 + \omega^2\tau^2} \tag{7.86b}$$

To proceed further we need *to* use the general expression for electromagnetic waves developed in Sect. 7.6. Recall Eq. 7.62:

$$n - ik = \sqrt{K_\mathrm{m}}\sqrt{K_\mathrm{e} - i\frac{\sigma}{\varepsilon_0\omega}} \quad \text{(Previous Eq. 7.62)}$$

and from the complex σ of Eq. 7.85, we can write:

$$n - ik = \sqrt{K_\mathrm{m}}\sqrt{K_\mathrm{e} - i\frac{\sigma_1 - i\sigma_2}{\varepsilon_0\omega}} \tag{7.87}$$

If we assume only free-electron optical interactions, the metal does not polarize under the wave, and $K_\mathrm{e} = 1$. In addition, for metals in the infrared, $K_\mathrm{m} = 1$. Thus,

$$n - ik = \sqrt{1 - i\frac{\sigma_1 - i\sigma_2}{\varepsilon_0\omega}} \tag{7.88a}$$

or

$$n - ik = \sqrt{1 - \frac{\sigma_2}{\varepsilon_0\omega} - i\frac{\sigma_1}{\varepsilon_0\omega}} \tag{7.88b}$$

It remains only to separate the real and imaginary parts of Eq. 7.26, which will yield two equations in n and k and thus give n and k in terms of the dc conductivity or and the relaxation time τ. Then we can use our expression for the reflectivity from Eq. 7.76 to generate R from n and k.

To carry out the algebra we use the identity

$$\sqrt{A + iB} = \sqrt{\frac{R + A}{2}} + i\sqrt{\frac{R - A}{2}}$$

Where $R = \sqrt{A^2 + B^2}$

Letting

$$A = 1 - \frac{\sigma_2}{\varepsilon_0 \omega}$$

and

$$B = -\frac{\sigma_1}{\varepsilon_0 \omega}$$

We have

$$2n^2 = \left(i - \frac{\sigma_2}{\varepsilon_0 \omega}\right) + \sqrt{\left(1 - \frac{\sigma_2}{\varepsilon_0 \omega}\right)^2 + \left(\frac{\sigma_1}{\varepsilon_0 \omega}\right)^2} \tag{7.89a}$$

$$2k^2 = -\left(1 - \frac{\sigma_2}{\varepsilon_0 \omega}\right) + \sqrt{\left(1 - \frac{\sigma_2}{\varepsilon_0 \omega}\right)^2 + \left(\frac{\sigma_1}{\varepsilon_0 \omega}\right)^2} \tag{7.89b}$$

Equations 7.89a and 7.89b, together with Eq. 7.76 give the reflectivity of R. Notice that R is a function solely of σ_1, σ_2 and ω. Look again at Eqs. 7.86a and 7.86b and note that σ_0 can be used to replace τ in the expression for σ_1, and σ_2 as:

$$\sigma_1 = \frac{\sigma_0}{1 + \omega^2 \dfrac{m^2 \sigma_0^2}{N^2 e^4}} \tag{7.90a}$$

$$\sigma_2 = \sigma_0^2 \frac{\frac{m}{Ne^2}\omega}{1 + \omega^2 \dfrac{m^2 \sigma_0^2}{N^2 e^4}} \tag{7.90b}$$

Equation 7.90a and 7.90b show that σ_1, and σ_2, and thus R, depend on frequency ω, constant m/N, and dc conductivity σ_0. Thus,

$$R = f(\omega, \sigma_0, m/N) \tag{7.91}$$

This means that we can use the dc conductivity to predict the reflectivity. Furthermore, if we know the temperature variation of σ_0, we can use this method to calculate R as a function of temperature. This is a useful result, because it is difficult to measure optical reflectivity as a function of temperature, whereas it is fairly easy to measure σ_0 vs. temperature.

A wealth of data on electrical conductivity has been amassed for most metals and alloys. Thus, the free-electron model is currently enjoying a great deal of attention as a way of providing reflectivity-vs-temperature information in the study of laser effects.

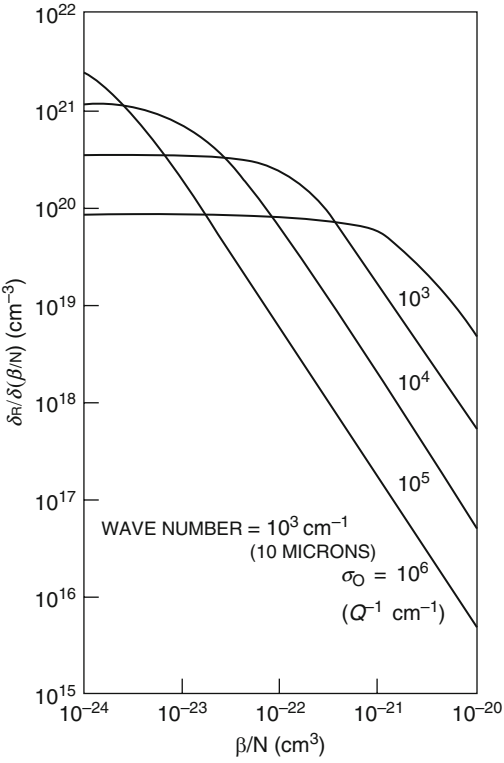


Fig. 7.18 Sensitivity of R to the parameter β/N [10]

There is, of course, one problem in using $\sigma_0(T)$ data to predict R, and that is the parameter m/N. It turns out that R is fairly insensitive to this parameter at infrared wavelengths. To see this we show here some numerical illustration. Define

$$\beta = m/m_0 \tag{7.92}$$

where m_0 is the free electron mass, and then the parameter β/N is equivalent to m/N.

Figure 7.18 shows a plot of $\partial R/\partial(\beta/N)$ as a function of β/N for $\lambda = 10\,\mu\text{m}$ and various values of σ_0. Here σ_0 is in units of reciprocal ohm-centimeters. Typical values are, for example, $\sigma_0 = 10^5\Omega^{-1}\,\text{cm}^{-1}$ and $\beta/N = 10^{-23}\,\text{cm}^3$ for aluminum. Then the value of $\partial R/\partial(\beta/N)$ is about $7.2 \times 10^{20}\text{cm}^{-3}$. If we take a 10 % error in β/N, we get

$$\frac{\partial R}{\partial(\beta/N)}\Delta(\beta/N) = 7.2 \times 10^0 \times 10^{-24}$$

$$\Delta R = 0.00072$$

Since for these values $R = 0.97366$, the change in R? is only about 0.1 %. We can obtain quite good predictions by the Drude–Lorentz model using the experimental

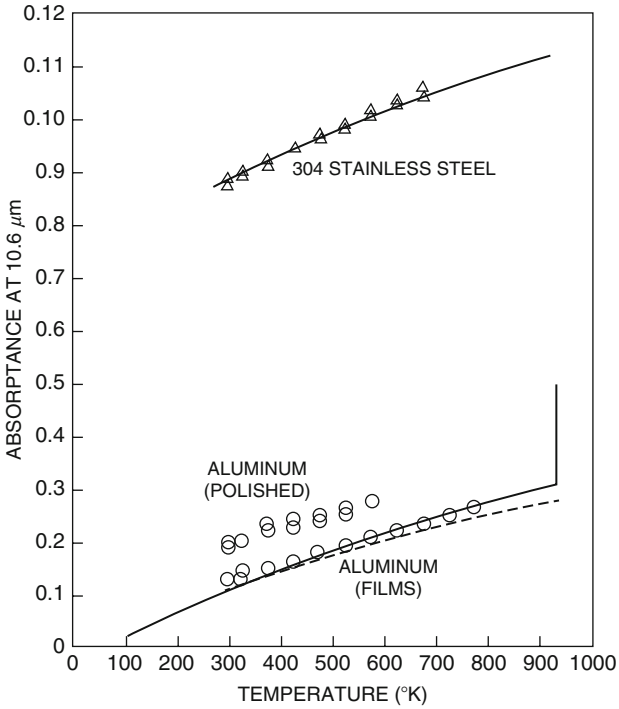


Fig. 7.19 Temperature dependence of the absorptance at 10.6 μ for aluminum and stainless steel

values of σ_0 and the most simple choice for β/N, namely one free electron for each valence electron per atom in the metal, and $\beta = 1$. For alloys, it is sufficient to choose the major constituent of the alloy. For example, with stainless steel we choose iron, or two electrons per atom, to compute N and hence β/N.

Figures 7.19 and 7.20 do show the predictions of the free-electrons theory for a variety of metals and some comparison to experimental data [10]. The abrupt change when the metal molts is caused by the abrupt change- in the 0 conductivity. Notice in the comparison to data that aluminum films give values closet to the theory. This is probably because they prevent the best surfaces. Defects, oxide layers, etc., tend to trap the incident radiation and cause they real surface to absorb more radiation than the ideal surface. These graphs are in terms of absorptance, which is the experimentally measured quantity, and, since metals are opaque, $A = 1 - R$, which is correct for specular reflection at normal incidence from an opaque substance.

Let us return to the expressionism for n and k to look at some limiting forms and thus show how these complete expressions reduce to simple relationships. Remember that R (Eq. 7.76) is determined by n and k (Equations sets 89a and 89b), which are in turn obtained from the dc conductivity and m/N (Equations sets 7.90a and 7.90b). The variation of n and k with wavelength is shown in Fig. 7.21 for a typical good conductor like aluminum or copper at room temperature. Note that at long

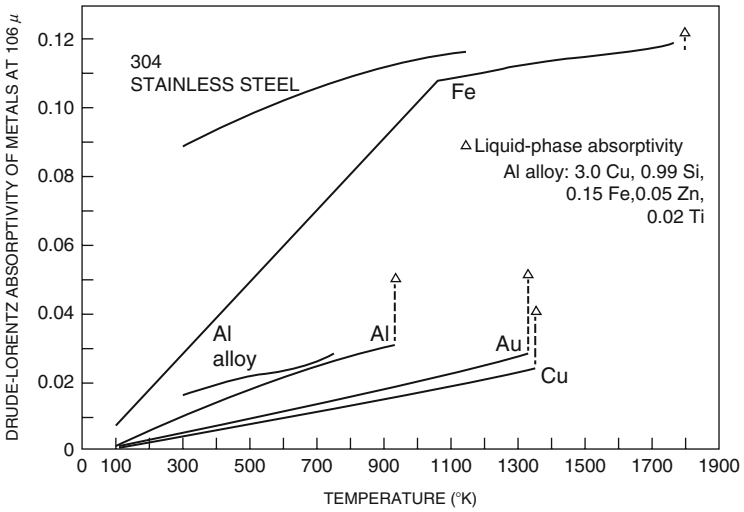


Fig. 7.20 Free-electro theory predictions of absorptivity of several metals at 10.6 μ. The *open symbols* indicate the molten state

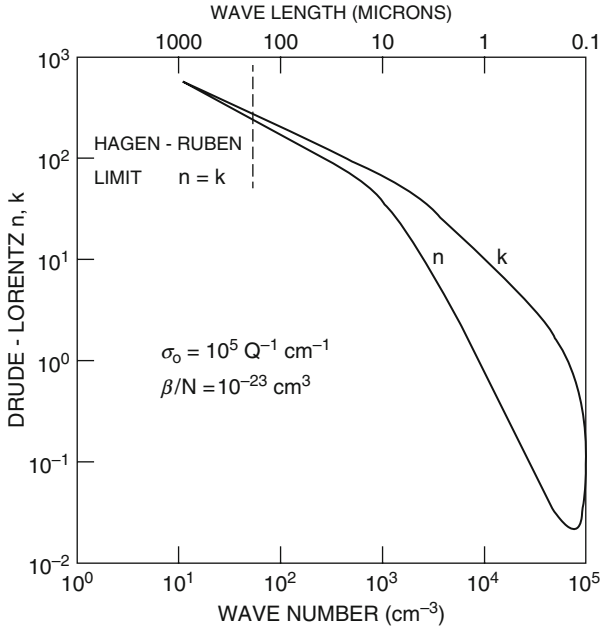


Fig. 7.21 *n* and *k* as functions of wavelength [10]

wavelengths $n = k$. We can derive this by using Equations sets 86a and 86b for σ_1 and σ_2 and noting that as $\omega \to 0$, $\sigma_1 \to \sigma_0$ and $\sigma_2 \to \sigma_0 \omega\tau$. By substituting these values into Equation sets of 89a and 89b, we can readily show that

$$n = k = \sqrt{\frac{\sigma_0}{2\varepsilon_0 \omega}} \tag{7.93}$$

This is called the Hagen–Rubens limit. Note that n is very large. Under these conditions algebra can be used to reduce Eq. 7.76 to the following form:

$$R = \frac{n-1}{n+1} \tag{7.94a}$$

or

$$R = 1 - \frac{2}{n} \tag{7.94b}$$

and Eq. 7.93 can be substitute for n to get:

$$R = 1 - 2\sqrt{\frac{2\omega\varepsilon_0}{\sigma_0}} \tag{7.95}$$

This is the Hagen–Rubens reflectivity.

We can also comment on the skin depth. We have, at long wavelength ($\omega \to 0$),

$$\delta = \frac{\lambda}{2\pi}\sqrt{\frac{2\varepsilon_0\omega}{\sigma_0}} \tag{7.96}$$

This can be rewritten as:

$$\delta = \sqrt{\frac{2}{\mu_0\sigma_0\omega}} \tag{7.97}$$

Equation 7.97 is the common expression for skin depth used at long wavelength.

Finally, we see from Fig. 7.22 that n and k reconverge at short wavelengths. This is called the plasma resonance. To see this, one must look at the behavior of n and k over a larger spectrum. We have already discussed the long-wavelength limiting behavior of n and k. This is the Hagen–Rubens region, where $n = k$. At short wavelengths, it is easy to show from Equations sets 7.90a and 7.90b that:

$$\sigma_1 \to \frac{N^2 e^4}{m^2 \sigma_0 \omega^2} \tag{7.98a}$$

$$\sigma_2 \to \frac{Ne^2}{m\omega} \tag{7.98b}$$

Thus, Eqs. 7.89a and 7.89b can be written, for large ω, as:

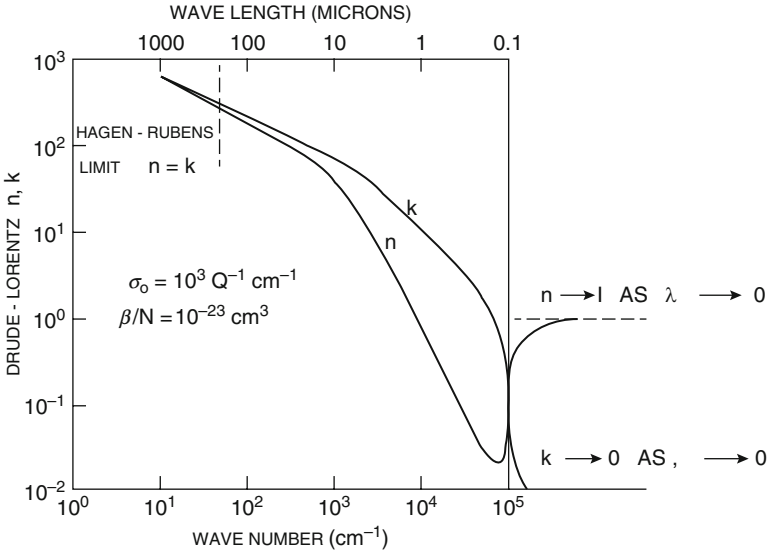


Fig. 7.22 n and k as functions of wavelength

$$n^2 = 1 - \frac{Ne^2}{\varepsilon_0 m\omega^2} \tag{7.99a}$$

$$k^2 = 0 \tag{7.99b}$$

Now the plasma frequency is usually defined from Eq. 7.99a by setting $n = 0$ to yield

$$\omega_p^2 = \frac{Ne^2}{\varepsilon_0 m} \tag{7.100a}$$

and thus

$$n^2 = 1 - \frac{\omega_p^2}{\omega^2} \tag{7.100b}$$

We see, then, that at very high frequencies the free-electron model predicts a transparent behavior $(k = 0)$ and the index of refraction approaches that of a vacuum. The transition to this transparent behavior takes place at the plasma frequency, and it is a fairly abrupt transition, as Fig. 7.22 shows. In fact, some texts call this transition the "ultraviolet catastrophe."

Note that at ω near ω_p Equations sets of 99a and 99b and 100a and 100b) are not valid. For these frequencies we must use the full expression. If we use again the

values of $\sigma_0 = 10^5 \Omega^{-1}\text{cm}^{-1}$ and $\beta/N = 10^{-23}\text{cm}^3$, which are appropriate to a good conductor like aluminum at room temperature, the reflectivity looks like Fig. 7.23. One can see that, in terms of the reflectivity, the transition is very abrupt, indeed.

The optical reflectivity of real metals is, as we have seen, in reasonable accord with the free-electron model at wavelengths in the infrared. The surface, however, must be nearly perfect for the predicted reflectivities to be achieved, and, of course, as the wavelengths approach the visible region band effects become important and the reflectivity shows rapid -Actuations with frequency. The absorptance of a practical metal surface is still largely an empirical matter. For high-power, continuous-wave radiation by a CO_2 laser, some data are available, but very little information on absorptivity as a function of surface temperature under these conditions is available.

Shown in Table 7.1 are room temperature absorptances for a few materials.

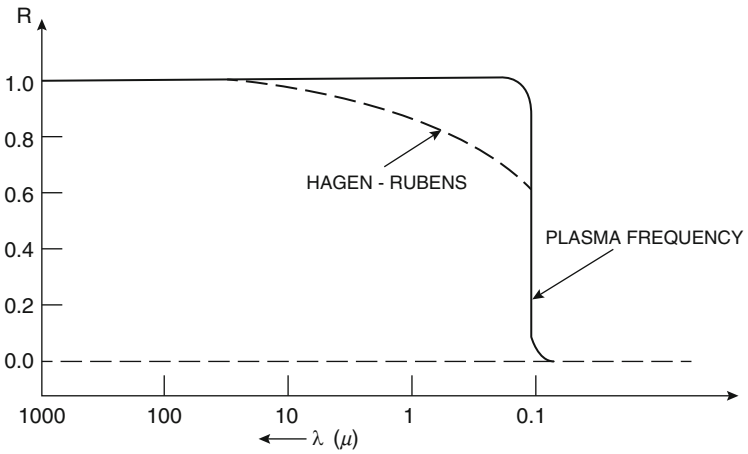


Fig. 7.23 *R* as a function of wavelength

Table 7.1 Room-temperature absorptance of aerospace metals and alloys at 10.6 μm for various surface conditions and at normal incidence

	Surface condition			
Metal or alloy	Ideal	Polished	As-received	Sandblasted
Al	0.013	0.03	0.040 ± 0.02	0.115 ± 0.015
Au	0.006	0.01	0.02	0.14
Cu	0.011	0.016		0.06
Ag	0.005	0.011		
2014 Al		0.033	0.07 ± 0.02	0.25
304 Stainless steel		0.11	0.4 ± 0.2	
Ti Alloy (6Al, 4 V)			0.65 ± 0.2	
Mg alloy Az-31B			0.06 ± 0.03	

Data on the reflectivity of a metal during actual irradiation by a laser beam is quite difficult to obtain, although this information is central to the problem of laser–material interaction. One classic experiment along these lines was carried out by Bonch-Bruevich, Imas, Romanov, Libenson, and Maitsev in Russia in 1967 [36]. They surrounded their specimens with a sphere to monitor the reflected radiation, as shown schematically in Fig. 7.24. The output of the photodetector is proportional to the reflectance of the specimen. Some of their results for steel and copper are shown in Fig. 7.25. The laser pulse (Nd: glass laser, 1.06 μm), with a peak power density of

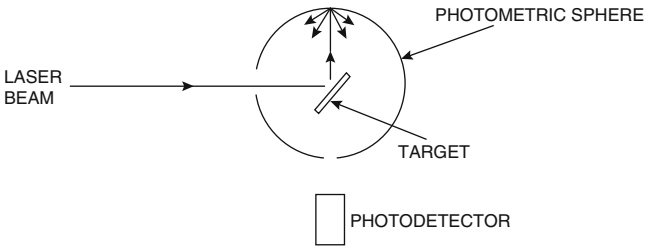


Fig. 7.24 Schematic representation of Bonch-Bruevich experiment

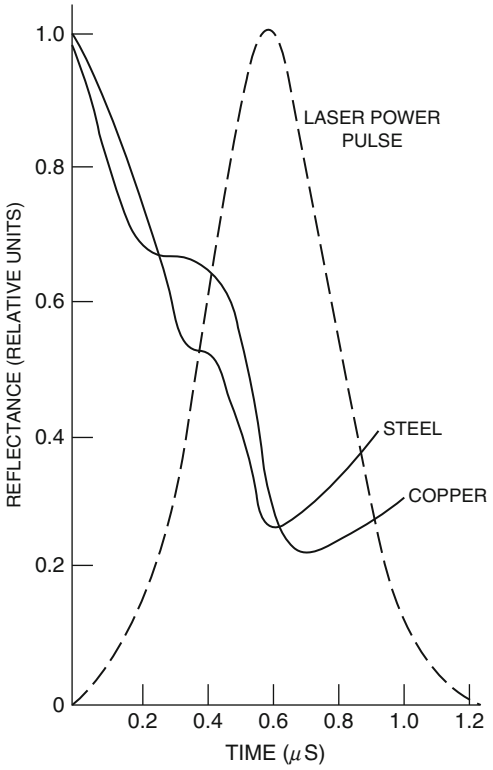


Fig. 7.25 Reflectance of steel and copper during irradiation by a laser beam [3]

the order of 10^8 W/cm^2, is shown as a broken line. As time passes, of course, the laser pulse heats the surface and the reflectance decreases. An especially interesting feature of these data is the shoulder. The author has suggested that this leveling off is associated with the surface reaching the melting point and pausing at that temperature while the thickness of the molten layer propagates into the solid. In short, order, however, the molten layer begins to heat up and the reflectance continues to decrease. As the power density of the laser pulse reaches its peak and begins to fall, the surface temperature can no longer be maintained, and as the surface cools the reflectance begins to increase again.

In calculation of the laser interacting with metals commonly used material properties are: (for energy balance analysis) density, heat capacity, specific heat ratio, heat conductivity, heat diffusivity, latent heat, melting point, vaporization point; (for stress and momentum analysis): viscosity, modulus of elasticity, shears modulus, Poison's ratio, stress-strain constitutive relation.

7.7 Mathematical of Laser Absorption in Metals

A complete understanding of laser interaction with materials is still a matter of trials and adjustments. The real physical processes of laser beam interaction (drilling, cutting, welding, or being used as a directed energy weapon application) with materials are very complex. Problem of laser interaction with materials presents many difficulties, both from modeling as well as from experimental sides. One would expect a reasonable description of the main phenomena occurring during laser interaction, but this is complicated because many of physical processes equally contribute to the development of conservation equations, producing draw back because of a great complexity of the equations to be solved. In most instances, this leads to formulation of a model needed to be solved numerically. A lack of pertinent experimental data to compare with, forces one to simplify some equations and use previous analytical and computational work done in this field [4].

The absorption coefficient, which can be derived from the material's dielectric function and conductivity, determines the absorption of light as a function of depth. However, the specific mechanisms by which the absorption occurs will depend on the type of material. In general, photons will couple into the available electronic or vibrational states in the material depending on the photon energy. In insulators and semiconductors, the absorption of laser light predominantly occurs through resonant excitations such as transitions of valence band electrons to the conduction band (interband transitions) or within bands (intersubband transitions) [20]. These excited electronic states can then transfer energy to lattice phonons. Photons with energy below the material's band gap will not be absorbed (unless there are other impurity or defect states to couple to or if there is multi-photon absorption). Such energies typically correspond to light frequencies below vacuum ultraviolet for insulators and below the visible to infrared spectrum for semiconductors. However,

resonant coupling to high-frequency optical phonons in the near-infrared region is possible in some cases [31].

As we discussed in previous sections, light impinging on a material surface can be reflected, transmitted, or absorbed. In reality, all three occur to some degree. In order for laser machining to be practical, the laser light must be absorbed by the material. To yield an efficient process, it is necessary to couple as much of the incident intensity to the work-piece as possible. This coupling efficiency is described by the sample absorptivity A (in some parts of the book this is also referred to as absorptance, absorption coefficient, or just absorption). The absorptivity is defined as the ratio between the absorbed energy and the incident energy. Absorptivity changes during the heating process and is a function of the sample's optical properties as well as the properties of the electromagnetic wave.

The mathematical model must be derived such that the parameters are easily handled. Once the model is verified by experiments, it can simulate the process and provide information such as heat-affected zone, transient temperature distribution, and cooling rates. Therefore, the model can reduce the experimentation by determining the effects of particular parameters beforehand.

This section presents the formulation of the general governing equation using the concept of the divergence of transport intensity as the net accumulation rate of energy per unit volume of the medium under consideration. Prediction of thermal effects produced by laser beam scanning the surface of an absorbing sample requires that three dimensional heat transfer equation be solved subject to finite size conditions of a sample. The geometry considered in this dissertation is illustrated in Fig. 7.26, which is finite in dimensions slab irradiated by a laser beam impinging on its surface subject to convective and radiative losses.

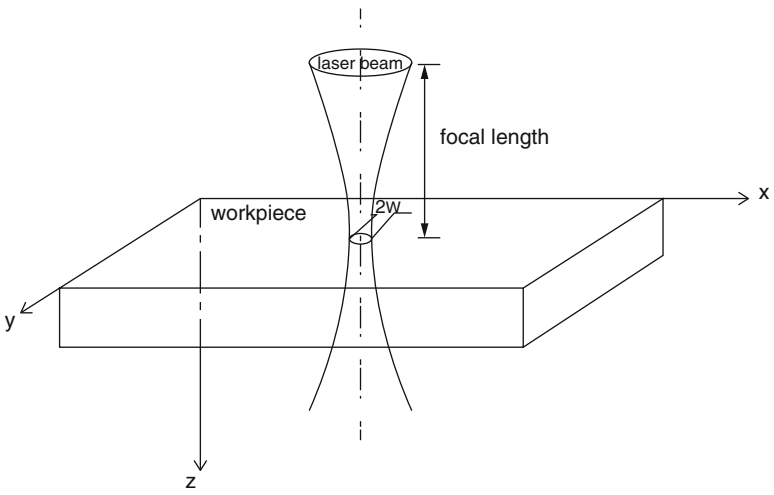


Fig. 7.26 Laser beam impinging on a finite size sample

As part of dealing with high-power laser interaction with materials and analysis of thermal response that is discussed in Sect. 7.8, we can define heat flux q_n as the magnitude of the in the n-direction (with n being x, y, z) one may introduce the vector sum, $\vec{q}$, of the directional fluxes as:

$$\vec{q} = q_x\hat{i} + q_y\hat{j} + q_z\hat{k} \tag{7.101}$$

Where $(\hat{i}, \hat{j}, \hat{k})$ are the unit vectors along each of the Cartesian coordinate directions.

Having defined a vector representation of the energy transport involved, one may write the expression for the net accumulation rate of thermal energy per unit volume as:

$$\nabla \cdot q + Q = \frac{\partial u}{\partial t} \tag{7.102}$$

where Q is a volumetric term accounting for the internal generation of heat u. The right hand side of Eq. 7.102 represents the total rate of change of the element's specific internal energy. Thus, Eq. 7.102 is a general statement of the first law of thermodynamics.

Using thermodynamic arguments, Eq. 7.102 can be expended further. If one considers sample under consideration to be a homogenous continuum composition, with more than one phase, than one can uniquely determine the state of any property of the continuum using two independent properties, specific energy and specific enthalpy, one can write [37].

$$du_{\mathrm{m}} = \left(\frac{\partial u_{\mathrm{m}}}{\partial \upsilon}\right)_T d\upsilon + c_\upsilon dT \tag{7.103}$$

$$dh_{\mathrm{m}} = \left(\frac{\partial h_{\mathrm{m}}}{\partial p}\right)_T dp + c_p dT \tag{7.104}$$

where

u_{m} = Specific internal energy
h_{m} = Specific enthalpy
υ = Volume
p = Pressure
c_υ = Specific heat at constant volume
c_p = Specific heat at constant pressure
T = Temperature distribution

Both the specific heat at constant pressure and volume are defined as below respectively:

$$c_p = \left(\frac{\partial u}{\partial T}\right)_p \tag{7.105}$$

$$c_v = \left(\frac{\partial u}{\partial T}\right)_v \tag{7.106}$$

For solids and incompressible fluids the specific volume is assumed constant. If one also neglects pressure changes in the given process, Eqs. 7.105 and 7.106 can be simplified to yield:

$$du_{\mathrm{m}} = c_v dT \tag{7.107}$$

$$dh_{\mathrm{m}} = c_p dT \tag{7.108}$$

Recalling the definition of specific enthalpy from classical thermodynamics as:

$$h_{\mathrm{m}} = u_{\mathrm{m}} + pv \tag{7.109}$$

One can combine Eqs. 7.105 and 7.103 to find that:

$$c_p = c_v = c \quad \text{for} \quad dv \approx dp \approx 0 \tag{7.110}$$

If pressure is allowed to vary, Eq. 7.109 still holds approximately for solids and incompressible fluids. Taking the relations in Eqs. 7.107 and 7.109 and multiplying by the density of the medium results in expressions in terms relative to the unit volume as opposed to the unit mass. Using Eq. 7.109, these new volumetric terms can be expressed as:

$$du = \rho \cdot c(T) dT \tag{7.111}$$

$$dh = \rho \cdot c(T) dT \tag{7.112}$$

and

$$h = u + \rho dT \tag{7.113}$$

Integrating both sides of Eq. 7.111 over proper variables and differentiating with respect to time yields

$$\frac{\partial u}{\partial t} = \rho \cdot \frac{\partial}{\partial t} \int_{T_0}^{T} c(T) dT \tag{7.114}$$

Where T_0 is the temperature at the beginning and T is the temperature at the end of the infinitesimal time step, dt. For the particular very infinitesimal time interval dt one can assume a constant specific heat, then Eq. 7.114 simplifies to:

$$\frac{\partial u}{\partial t} = \rho \cdot c \frac{\partial T}{\partial t} \tag{7.115}$$

since the initial temperature is constant with respect to time. Substituting Eq. 7.115 into Eq. 7.102 yields the heat diffusion equation as:

$$-\nabla q + Q = \rho \cdot c \frac{\partial T}{\partial t} \tag{7.116}$$

Solution of Eq. 7.116 for the time-dependent field, $T(x, y, z, t)$, requires the use of a constitutive equation relating the temperature to heat flow. For conductive heart transfer equation this relation is Fourier's law of conduction. It states that the heat flux in a direction n is proportional to the temperature gradient in that direction. Mathematically this is expressed as

$$q_n \hat{n} = -k \cdot \frac{\partial T}{\partial n} \hat{n} \tag{7.117}$$

where k is thermal conductivity and q_n is the heat flux in the n-direction. Negative sign is necessary to satisfy the second law of thermodynamics.

In the analysis of the Fourier heat conduction model, the heat flux through a given plane is considered as being a function of the spatial temperature gradient at that plane.

This depends upon the assumption that the temperature gradient remains almost constant between two successive and closely spaced planes. However, the distance between these planes is finite, therefore, error occurs when high-order terms, which are neglected, become important at high-power laser intensities. The heat flux through a given plane depends on the electron energy distribution through the material, therefore, the material cannot be considered as a homogeneous continuum when one is analyzing very short pulses (shorter then a picosecond) at intermolecular level (distances less than 0.1 μm).

Substituting Eq. 7.117 into Eq. 7.116 yields

$$-\nabla \cdot (k \nabla T) + Q = \rho \cdot c \frac{\partial T}{\partial t} \tag{7.118}$$

which is the general form of the governing differential equation for isobaric thermal conduction in a homogenous solid or incompressible fluid.

In laser micromachining or its interaction with materials surface, the internal energy generation, Q, is commonly thought of as the rate of laser energy absorbed per unit volume in the irradiated medium. In case of metals, this absorption occurs

at beginning in a very thin layer at the surface of the workpiece or target of interest and, for many practical cases can be considered as a boundary condition to Eq. 7.118. This absorption process is calculated by Beer–Lambert's law, via the following relation as:

$$I(z) = \varepsilon \cdot I_0(0)|_{\substack{z=0 \\ \text{material surface}}} \exp(-\mu \cdot z) \tag{7.119}$$

where $I(z)$ is the intensity of the incident radiation at a given distance, z, into the absorbing medium from the irradiated surface, ε is the surface emissivity of the medium and μ is the absorption coefficient of the material, measure of the absorption of radiation propagating through the medium. Using the terminology used in Eq. 7.102, differentiating Eq. 7.117 with respect to the direction of propagation of the laser beam yields the volumetric extraction rate of energy absorbed by the differential element. Since the accumulation rate is negative of the extraction rate for a given volume, the necessary volumetric term for use in Eq. 7.119, is given by:

$$Q = -\frac{\partial I(z)}{\partial z} = (\varepsilon\mu) I_0(0)|_{\substack{z=0 \\ \text{material surface}}} \exp(-\mu \cdot z) \tag{7.120}$$

or for simplicity we write Eq. 7.111 as:

$$Q = -\frac{\partial I(z)}{\partial z} = (\varepsilon\mu) I_0(0) \exp(-\mu \cdot z) \tag{7.121}$$

Using the Drude–Zener theory [38], Eq. 7.121 leads to the following expression for $Q(x, y, z, t)$:

$$Q(x, y, z, t) = A\exp(-\beta t)[\mu I_0(x, y, z, t)\exp(-\mu z)] \tag{7.122}$$

where A is the surface absorptivity, β is the pulse parameter, and $I_0(x, y, z, t)|_{z=0}$ is laser radiation intensity at the material surface ($z = 0$).

In Eq. 7.122, absorption coefficient of the material ì measuring the absorption of radiation propagating through the medium is considered constant at this time, but based on a number of publications ì should be a function of temperature and axial position of the irradiated laser beam in respect to the material under consideration. At this time let us modify Eq. 7.122 to be

$$Q(x, y, z, t) = A\exp(-\beta t)[\mu(T, z) I_0(x, y, 0, t)\exp\{-\mu(T, z) \cdot z\}] \tag{7.123}$$

and function $\mu(T, z)$ will be defined as bulk absorption coefficient at distance z from metal surface. Here for purpose of our analysis and discussion we keep μ constant, but based on a number of publications μ should be a function of temperature and

axial position of the irradiated laser beam in respect to the material under consideration

Incorporation Eq. 7.122 into 7.118, for laser pulse with a position-dependent intensity, for this form of pulse input, the Fourier differential Eq. 7.118 can be rewritten a in very general form as:

$$\rho c(T)\frac{\partial T}{\partial t} = \nabla\left[k(T)\nabla T\left(x,y,z,t\right]\right. + Q(x,y,z,t) \tag{7.124}$$

where $c(T)$ is the density of the material of the workpiece, is the temperature dependent specific heat of the material, $k(T)$ is the temperature dependent thermal conductivity, $T(x,y,z,t)$ is the resulting three-dimensional time dependent temperature distribution in the material t is time and we can assume for the purpose of any boundary condition that T_0 is the initial temperature, and x,y,z are the spatial Cartesian coordinates, while $Q(x,y,z,t)$ is the rate at which heat is supplied to the solid per unit time per unit volume, depends on the laser pulse parameter and physical and optical properties of materials irradiated. Note that both A and μ are functions of temperature and wavelength of the incident radiation as shown in the later sections of this chapter.

Solid or liquid evaporates at any temperature greater than 0 K. The evaporation rates strongly depend on the surface temperature T_s. Equation 7.124 considers heat diffusion into material only through conduction. Based on experimental and theoretical evidence, evaporation takes place, so one has to consider moving elements of vapor and liquid inside the material during laser beam interaction.

Considering this fact, the general governing differential equation allowing a phase change process can be written as:

$$\rho(T)c(T)\frac{\partial T}{\partial t} = \nabla(k\nabla T) + \rho(T)c(T)V_n(T)\nabla(T) + Q \tag{7.125}$$

where V_n is the normal component of the evaporation front or melt front velocity (recession velocity)

Let us define velocity, V_n. From literature [39], the interface moves into depth of the material at the speed defined as:

$$V(T_s) = \frac{1-c}{\rho}\sqrt{\frac{M}{2\pi k_B T_s}p_{sat}(T_s)} \tag{7.126}$$

where c is recondensation factor which is usually taken as 0.18 and $p_{sat}(T_s)$ is saturated vapor pressure defined from Clapeyron–Clausius Equation as:

$$p_{sat}(T_s) = \varphi_0 \exp\left(-\frac{E_a}{k_B T_s}\right) \tag{7.127}$$

and k_B is Boltzmann's constant, φ_0 is pre-exponential factor, and E_a is an activation evaporation per atom.

Note that Eq. 7.126 is a pressure dependent function, which has to be calculated simultaneously with temperature dependent velocity. Thus, let us derive a velocity as a function of temperature and the latent heat of phase transition, which is also temperature dependent.

The rate of change of latent heat with temperature can be expressed as [40].

$$\frac{\partial H}{\partial T} = \frac{H}{T} + \left(c_{pv} - c_{pl}\right) - \frac{H}{\upsilon_v - \upsilon_l}\left[\left(\frac{\partial \upsilon_v}{\partial T}\right)_p - \left(\frac{\partial \upsilon_l}{\partial T}\right)_p\right] \tag{7.128}$$

where c_{pv} and c_{pl} are specific heats at constant pressure for vapor and liquid states, respectively, and υ_v and υ_l are specific volumes for vapor and liquid states.

One has to realize that integration of the latent heat over the temperature ranging from 0 to critical temperature T_c is difficult, because we do not have enough information about values for the latent heat especially at extreme value of the range under consideration. It is safe to assume [41], that little inaccuracy is involved in taking the room temperature latent heat as the latent heat at absolute zero, because, by knowing that $\upsilon_v << \upsilon_l$ and $\left(\frac{\partial \upsilon_v}{\partial T}\right)_p >> \left(\frac{\partial \upsilon_l}{\partial T}\right)_p$, i.e., the specific volume of gas, υ_v is much greater than the condensed liquid, υ_l and its rate of change with temperature at constant pressure is corresponding greater. Thus, Δc_p is extremely small for temperature up to room temperature T_a.

According to Maxwell's law the function of velocity distribution of molecules can be defined as [42].

$$f(V_n)dV_n = \sqrt{\frac{M}{2\pi k_B T_s}\exp\left(\frac{MV_n^2}{2\pi k_B T_s}\right)dV_n} \tag{7.129}$$

where V_n is the velocity in the direction normal to the surface, and the other parameters were defined earlier. Using vernacular terms, function $f(V_n)dV_n$ is ratio of number of atoms with velocity V_n to $V_n + dV_n$ per unit volume, to the total number of atoms per unit volume. Only those molecules whose velocity is greater than V_{min} obtained from Equation given by [42] as:

$$\frac{1}{2}MV_{min}^2 = H(T) \tag{7.130}$$

will escape from the retaining potential, where V_{min} lies in the z direction. If n is the number of atoms per unit volume then the number of atoms with velocities V_n to $V_n + dV_n$ per unit volume is $nf(V_n)dV_n$ and the number of atoms with these velocities passing a unit area per unit time is $nf(V_n)V_n dV_n$

Assuming that all the atoms for which , do not return to their equilibrium position under evaporation circumstances and if $V_n > V_{min}$ and if we define as the number of atoms evaporated per unit time per unit area, then we can write:

$$N_G = \int_{V_{min}}^{\infty} [nf(V_n)V_n]dV_n$$
$$= n\sqrt{\frac{M}{2\pi k_B T_s}}\int_{V_{min}}^{\infty}\left[\exp\left(\frac{MV_n^2}{2\pi k_B T_s}\right)V_n\right]dV_n \tag{7.131}$$

After integration and substitution of Eq. 7.130 into 7.131, we obtain:

$$N_G = n\sqrt{\frac{M}{2\pi k_B T_s}}\exp\left(\frac{H(T)}{k_B T_s}\right)V_n \tag{7.132}$$

If atoms are equally spaced within the lattice a surface layer would consist of $n^{2/3}$ with an evaporation time $n^{2/3}N_G$. The average velocity of the surface, V_n would be:

$$V_n = \frac{1}{n^{1/3}}\frac{N_G}{n^{2/3}} \tag{7.133}$$

$$V_n = \sqrt{\frac{k_B T}{2\pi M}}\exp\left(-\frac{H(T)}{k_B T}\right) \tag{7.134}$$

Where $H(T)$ is the *Latent Heat*.

Considering Eq. 7.134, the general governing differential equation define in Eq. 7.135 below, with phase change processes at the surface of material interaction with high-power laser irradiation could be written as:

$$\rho(T)c(T)\frac{\partial T}{\partial t} = \nabla(k\nabla T) + \rho(T)c(T)\sqrt{\frac{k_B T}{2\pi M}}\exp\left(-\frac{H(T)}{k_B T}\right)\nabla T + Q \tag{7.135}$$

Note that latent heat $H(T)$, in Eqs. 7.134 and 7.135, is based on some theoretical and experimental evidence is temperature dependent and the final form of it is going to be described in later section.

In order to solve Eq. 7.134, appropriate boundary conditions should be applied. Section 7.8.1 discusses formulation of the appropriate boundary conditions.

7.8 Material and Thermal Response

Metal processing with lasers has reached a high level of maturity and acceptance in industry. It is used for cutting, drilling, welding, forming, engraving, marking, hardening, and various forms of surface treatment of metals in a broad spectrum

of modern industries, including the automotive and aerospace industries, the shipbuilding industry, the microelectronics industry and the medical instrument industry to name a few and now its application as a directed energy weapon both based on ground and space usages.

The details of the material response will depend on the particular material system and the laser processing conditions. As was mentioned earlier, if laser induced excitation rates are slow compared to the thermalization time, then the process is denoted as photothermal, and one can consider the absorbed laser energy as being directly transformed into heat. In this case, the material response will be a function of the local material heating and cooling rates, maximum temperatures reached, and temperature gradients, all of which can be determined from the solution to the heat equation for the given irradiation conditions. Because material heating rates can be so extreme, reaching as high as 109 K/s for nanosecond (ns) pulses and even higher for femtosecond lasers, significant changes to the material can occur.

One of the most important effects of intense laser irradiation is the conversion of the optical energy in the beam into thermal energy in the material. This is the basis of many applications of lasers, such as welding and cutting. We shall summarize here this thermal response. It is basically a classical problem, namely heat flow. In the usual manner, we shall seek solutions to the equation which governs the flow of heat, namely

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(K \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(K \frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(K \frac{\partial T}{\partial z}\right) + A \tag{7.136}$$

We use here ρ for the density, c for the specific heat, T for temperature, I for time, and K for thermal conductivity. A is the heat produced per unit volume per unit of time. In Eq. 7.136, ρ, c, and K are considered functions of both position and temperature, and A is a function of both position and time. In effect, the equation is a simple statement that the rate at which heat accumulates in an elemental volume $dx\,dy\,dz$ is equal to the net flow of heat across the faces of that volume plus the rate at which heat is produced within the volume.

Thus, thermal response studies consist essentially of two parts. First, one needs to know the rate and source of production of heat by the laser, which yields A. Then one solves Eq. 7.136 subject to the boundary conditions of the situation of interest. This can be a very elaborate task and frequently can be done only with the aid of a computer.

There is a great deal of effort among workers in the field of laser effects to develop an all-inclusive computer program to solve Eq. 7.136 for every possible situation. However, the solution to Eq. 7.136 can be no better than the knowledge of A, and, as we see in later sections, it is often very difficult to establish A with any precision in a laser material interaction situation.

7.8.1 Boundary Conditions

Boundary conditions are important part of a model. They influence the programming and calculation results greatly. Specifying the suitable boundary conditions is the basis for successful computation. Three kinds of boundary conditions are typically encountered in heat conduction analysis. These are:

1. Given the boundary temperature, T.
2. Given the boundary heat flux, q.
3. A boundary heat flux balance relation.

If the boundary temperature is given, there is no particular difficulty in modeling, one needs only to specify the value of the boundary grids to the specified temperature.

The magnitude of heat flux due to convection to the ambient from the sample surface is expressed using Newtonian law of cooling

$$q_{\text{conv}} = h_{\text{c}}(T_{\text{s}} - T_{\text{amb}}) \tag{7.137}$$

where h_{c} is convection heat transfer coefficient, T_{s} is the surface temperature, and T_{amb} is the temperature of ambient or surrounding.

In order to determine h_{c}, the characteristic length L of the workpiece should be:

$$L = \frac{\text{Area}}{\text{Ob}} \tag{7.138}$$

where Area is the area of the surface, and Ob is the perimeter of the workpiece. Then, the Nusselt number for the horizontal plate is [43]:

$$N_{\text{u}} = 0.27(R_{\text{a}})^{0.25} \tag{7.139}$$

where the Rayleigh number R_{a} is:

$$R_{\text{a}} = \frac{g\beta}{\kappa\gamma} L^3 (T - T_{\text{amb}}) \tag{7.140}$$

In Eq. 7.140, g is the gravitational acceleration, β is the coefficient of the volumetric thermal expansion, κ is the thermal diffusivity, γ is the kinematic viscosity, and other parameters are as defined for Eq. 7.137. The convective heat transfer coefficient h_{c} can be calculated as:

$$h_{\text{c}} = N_{\text{u}} \frac{k_{\text{air}}}{L} \tag{7.141}$$

where k_{air} is the thermal conductivity of the air surrounding the workpiece. Note that in case of high-power laser as directed energy weapon and its application as

Ground Based Laser (GBL) or Air Born Laser (ABL), we need to consider the above discussion. Radiation to the ambient is expressed using the relation:

$$q_{\text{rad}} = \sigma\varepsilon\left(T_{\text{s}}^4 - T_{\text{amb}}^4\right) \tag{7.142}$$

where q_{rad} is the magnitude of radiation flux, ε is the emissivity of the material, and σ is the Stefan–Boltzmann constant. Note that convection is a linear function of temperature, whereas radiation is a nonlinear, due to its dependence on the difference of the fourth powers of the surface and ambient temperature.

The magnitude of convection and conduction in the overall transport of heat can be evaluated from the value of the Péclet number, P_{e}, which is defined by

$$P_{\text{e}} = \frac{uc_p L_{\text{r}}}{k} \tag{7.143}$$

Where u is velocity, L_{r} is the characteristic length, taken as the pool radius at the top surface of the weld pool and the other parameters as defined earlier.

Heat transported by a combination of convection and conduction mechanism is observed in the weld pool in laser micro-welding applications. When Péclet number is less than 1, the heat transport within the weld pool, occurs primary by conduction. When Péclet number is much higher than 1, then the convective heat transport is the main mechanism of heat transfer in the material.

If one is using shielding gas in laser micromachining then heat transfer coefficient h_{c} required in Eq. 7.141, is calculated from Mazumder and Steen [44] and Gordon and Cobonpue [45] for case of a vertically impinging jet

$$h_{\text{c}} = 13\text{Re}^{0.3}\text{Pr}^{0.35}k_{\text{g}}\frac{1}{B} \tag{7.144}$$

where B is the jet plate distance, is the Reynolds number at jet exit, is the Prandtl number for gas, and k_{g} is thermal conductivity of the gas.

Performing an energy balance on a boundary where both convection and radiation losses occur, one can relate the flux conducted to the interface to the convection and radiation losses by the expression

$$q_{\text{n}} = q_{\text{conv}} + q_{\text{rad}} \tag{7.145}$$

Substituting for q_{n} from Eq. 7.117, for q_{conv} from Eq. 7.137 and finally for q_{rad} from Eq. 7.142, into Eq. 7.145, we get

$$-k\left(\frac{\partial T}{\partial n}\right)_{\text{s}} = h_{\text{c}}(T_{\text{s}} - T_{\infty}) + \sigma\varepsilon\left(T_{\text{s}}^4 - T_{\infty}^4\right) \tag{7.146}$$

where the term on the left-hand side of the Eq. 7.146 represents the magnitude of the heat conducted normal to the boundary surface. Equation 7.146 should be

applied to all exposed boundary of the finite sample under consideration. Setting the right-hand side of Eq. 7.146 to zero accommodates insulated boundaries.

In laser micro-drilling or interaction with materials, Eq. 7.146 should be applied in the region of the newly vaporized material (hole area) and an appropriate heat transfer coefficient should be used. Since the question of modeling of the micro-drilling phenomena is associated with creating the hole formation has not yet been solved, then to chose an appropriate heat transfer coefficient is an educated guess and open to debate.

For the case when phase change transition takes place, the boundary condition at the liquid–vapor interface is

$$-k\left.\frac{\partial T}{\partial z}\right|_z + \rho v_{dv} H_v = (1-R)I_0 \tag{7.147}$$

where k is thermal conductivity of the solid or liquid phase, $\left.\frac{\partial T}{\partial z}\right|_z$ is the temperature gradient at the surface along the normal (z-axis), ρ is the density of the solid or liquid phase, v_{dv} is the component of boundary velocity due to evaporation, R is the reflectivity for the laser wavelength, and I_0 is the intensity of the laser beam at the surface. For surface temperatures less than approximately half of the critical temperature, the energy of evaporation per atom U can be assumed to be constant. Then the component of the boundary velocity due to evaporation v_{dv} was defined by Eq. 7.134, and also defined in similar fashion by Niedrig and Bostonjglo [46].

$$v_{dv} = V_0 \exp\left(\frac{-U}{T_s}\right) \tag{7.148}$$

where V_0 is a coefficient of the order of magnitude of the sound velocity, T_s is the surface temperature, and U is the energy of evaporation per atom defines as:

$$U = \frac{MH_v}{N_a k_B} \tag{7.149}$$

Where H_v is the latent heat of vaporization (per unit mass), k_B is Boltzmann's constant, M is the atomic mass, and N_a is Avogadro's number. Velocity V_n, defined in Eq. 7.138, is equal to v_{dv}.

The classic Stefan boundary condition is applied to the solid–liquid boundary (melting front $z = z_m$) as:

$$\rho H_f v_m = k_s \left.\frac{\partial T_s}{\partial z}\right|_{z=z_m} - k_l \left.\frac{\partial T_l}{\partial z}\right|_{z=z_m} \tag{7.150}$$

where H_f is the latent heat of fusion, v_m is the melting front velocity, and subscripts "S" and "l" represent "solid" and "liquid," respectively. The Stefan boundary condition assumes an instant transition from solid to liquid at the melting

temperature Tm and does not allow superheating at the melting front. The approximation is adequate for the slow velocities of melt front propagation typical of laser welding and drilling, where the melting kinetics can be disregarded.

To understand how the boundary conditions were defined, let us first comprehend material removal process together with energy transport in multiple phase transitions, which is going to be discussed later on. To establish effects of convective heat transfer coefficient due to evaporation recoil generated melt flow Semak et al. [47] performed two simulations. In the first one recoil pressure and related melt flow were disregarded and in the second case recoil pressure and melt flow were accounted for. Those observations they supported by calculating the melt surface temperatures at the axis of the laser beam for different values of absorbed intensity I_0 presented in Fig. 7.27. When recoil pressure and melt flow were included, the steady-state values of temperature were reached faster, and the maximum temperature were lower and the cooling rates much higher, than for the case where recoil pressure and flow were neglected. Analyzing, Fig. 7.27, one can calculate that ignoring recoil pressure and associated convective heat transfer results in 1-5 % error in calculations of surface temperature in the center of the laser beam.

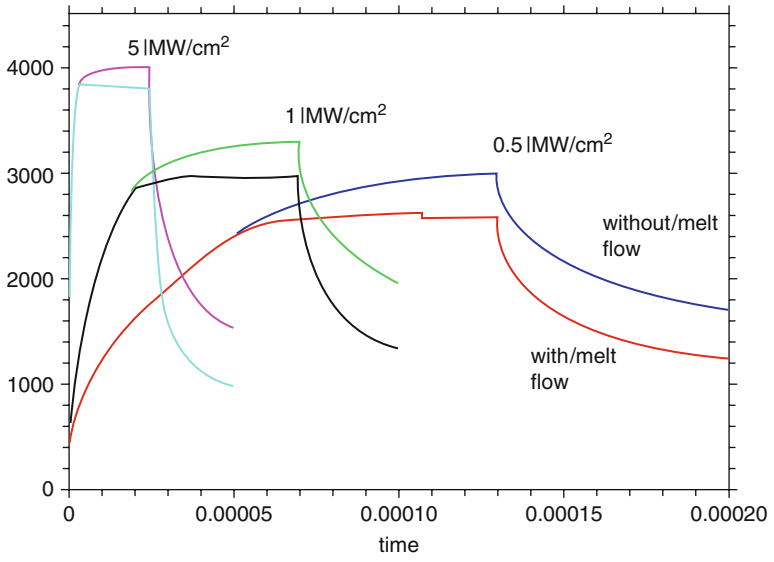


Fig. 7.27 Calculated temperature of iron surface at the beam axis for the cases without (*top curves*) and with (*bottom curves*) melt flow, for the different maximum absorbed intensity value and different laser pulse durations: 130 is (0.5 MW cm^2), 70 is (1 MW cm^2), and 25 is (5 MW cm^2) [47]

7.9 Solutions of Governing Equation

This section considers the solution of the General Governing Eq. 7.136 subject to its related boundary conditions, by means of exact and numerical methods. This section consists of a discussion of the approximate solution of the problem using mathematical tools and is divided into four subsections parts. First four sections Sects. 7.9.1.1, 7.9.1.2, 7.9.1.3, and 7.9.1.4 present an analytical method of the solution of the general governing equation using the Fourier theory. Section 7.9.1.5 considers interaction of laser energy with materials using very short laser pulses and introduces electron–phonon theory approach to solve the heat transfer problem of the interaction of ultra short pulses with the matter. Section 7.10 consists of a discussion of the approximate solution of the problem using the Finite Difference Method (FDM) and Finite Element Method (FEM), and presents the computer solutions developed

7.9.1 Analytical Methods

There have been many analytical solution published for various special cases of the Fourier heat transfer equation. Section 7.9.1.1 illustrates analytical solution for heating case without phase change in case of simple infinite solid, by laser defined as a step function (constant heating). Section 7.9.1.2 solves the problem of similar as Sect. 7.9.1.1, but considering the Slab of Finite Thickness. Section 7.9.1.3 deals with Analytical Solution of Heat Transfer Equation, with Spatial Dependent Laser Pulse Heating. Next Sect. 7.9.1.4 considers analytical solution to the governing heat transfer equation for laser pulse with a time-dependent Gaussian pulse heating. Section 7.9.1.5 considers analytical solution of heat transfer equation with time dependent Gaussian laser pulse heating with convective boundary conditions. Section 7.9.1.6 describes heating analysis with time dependent pulse intensity and where evaporation is considered as the exclusive phenomenon taking place during the ablation process. Section 7.9.1.7 presents the heating analysis with pulsed laser heating process by considering both Fourier conduction and electron–phonon kinetic theory approaches.

7.9.1.1 No Phase Change: Simple Infinite Solid

Before we go through such details, let us consider the simplest situation of high-power laser interaction with metal where we have No Phase Change and Semi-Infinite Solid case. Let the laser beam be perfectly uniform over an extremely large area, so that we have a one-dimensional situation. Assume also that the material parameters are temperature-independent and that the solid is uniform and isotropic and of semi-infinite extent (see Fig. 7.28). Finally, assume that there is no phase

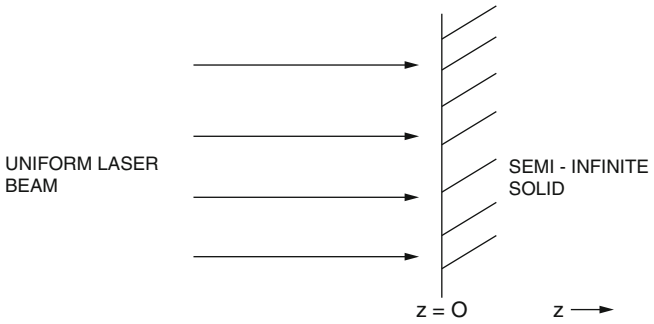


Fig. 7.28 Uniform irradiation of a semi-infinite solid

change; the rate at which energy enters the material is not sufficient to induce melting or vaporization.

First rewrite Eq. 7.136 in one-dimensional and in z-direction, using the fact that ρ, c and K are constant as:

$$\frac{\partial^2 T(z,t)}{\partial z^2} - \frac{1}{\kappa}\frac{\partial T(z,t)}{\partial t} = -\frac{A}{K} \tag{7.151}$$

Here we have introduced $\kappa = K/(\rho c)$, which is the thermal diffusivity. Let us adopt the convention that T is measured with respect. to the initial (or ambient) temperature of the material. This is possible because Eqs. 7.136 and 7.151 define T only to within an additive constant. Then we have as a boundary condition that $T \to 0$ as $z \to \infty$. The boundary condition on the front face $(z = 0)$ depend on what we assume for radiative and convective losses. It can be shown that, for most cases of interest, the rate at which the laser creates heat at the interface is overwhelmingly larger than convective and radiation losses, so we ignore them for the present calculation. Thus, the boundary condition is that there is no heat flux at $z = 0$, that is,

$$K\frac{\partial T}{\partial z}\bigg|_{z=0} = 0$$

Now consider A. Denote by I the power density of the laser radiation at the surface; the dimensions of I are power per unit area. The power density of the radiation transmitted to the surface is $I(1 - R)$. Then the power density as a function of z is:

$$F = (1 - R)Ie^{-4\pi kz/\lambda} \tag{7.152}$$

This follows from the fact that the energy in the electromagnetic wave goes as E^2. Now to get the power transferred per unit volume, consider elemental volumes of length dz and unit area:

$$A = -\frac{\partial F}{\partial z} = (1-R)Ie^{-4\pi kz/\lambda} \tag{7.153}$$

The minus sign appears because $\partial F/\partial z$ is the power per unit volume lost by the radiation and A is the power per unit volume absorbed by the material. Finally, we define the absorption coefficient

$$\alpha = \frac{4\pi k}{\lambda} \tag{7.154}$$

which is, of course $1/\delta$, the skin depth. Thus, Eq. 7.153 takes the following form:

$$A = (1-R)I(t)\alpha e^{\alpha z} \tag{7.155}$$

where we have included the possibility of I varying with time. Therefore the equation need to be solved is:

$$\frac{\partial^2 T}{\partial z^2} - \frac{1}{\kappa}\frac{\partial T}{\partial t} = -\frac{(1-R)I(t)\alpha e^{-\alpha z}}{K} \tag{7.156}$$

In keeping with our assumption of temperature-independent thermal parameters, we assume further that R is independent of temperature. Equation 7.155 is valid for temperature-dependent R and can be used to give $A(z,t,T)$.

For metals, α is a fairly large number. As we see in Sect. 7.6, k is of the order of 100 at $\lambda = 10\,\mu\text{m}$, so that α is of the order of $10^6\,\text{cm}^{-1}$. Hence, the absorption occurs in a very narrow layer at the surface. It then becomes more convenient to seek solutions for the homogeneous part of Eq. 7.156, as:

$$\frac{\partial^2 T}{\partial z^2} - \frac{1}{\kappa}\frac{\partial T}{\partial t} = 0 \tag{7.157}$$

subject to the boundary conditions that $T = 0$ at $z = \infty$, but with a specific flux into the surface at $z = \infty$, i.e.,

$$-K\left.\frac{\partial T}{\partial z}\right|_{z=0} = (1-R)I(t)$$

or, with the definition

$$F(t) = (1-R)I(t) \tag{7.158}$$

$$-K\left.\frac{\partial T}{\partial z}\right|_{z=0} = F(t) \tag{7.159}$$

First examine the case of $F = F_0$, a constant. This is appropriate to irradiation by a continuous laser, given temperature-independent material properties. We note here only the solution, for many excellent texts on heat conduction can be consulted for the details [3]. See also Example 5 in Appendix H. Also here we show detailed solution of this problem here as well.

Temperature in a Semi-infinite Solid

Before we solve this problem, we do couple of simple problem to establish a foundation and refresher for those readers that have out of these fundamental mathematic problem as well.

Problem 1: Let Γ denote the curved problem BJPKQLA of the Bromwich contour Fig. (a) with equation $s = Re^{i\theta}, \theta_0 \leq \theta \leq 2\pi - \theta_0$, i.e., Γ is the arc of a circle of radius R with center at O. Suppose that on Γ we have

$$|f(s)| < \frac{M}{R^k}$$

where $k > 0$ and M are constant. Show that

$$\lim_{R\to\infty} \int_{\Gamma} e^{st} f(s) ds = 0$$

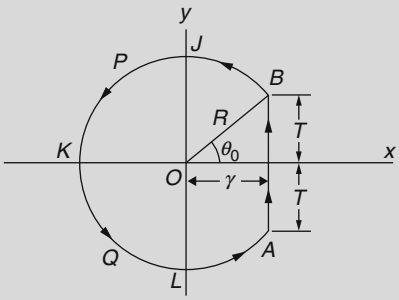


Fig. a

If $\Gamma_1, \Gamma_2, \Gamma_3$, and Γ_4 represent arcs of BJ, JPK, KQL, and LA, respectively, we have:

$$\int_{\Gamma} e^{st} f(s) ds = \int_{\Gamma_1} e^{st} f(s) ds + \int_{\Gamma_2} e^{st} f(s) ds + \int_{\Gamma_3} e^{st} f(s) ds + \int_{\Gamma_4} e^{st} f(s) ds$$

(continued)

Then if we can show that each of the integrals on the right hand side approaches zero as $R \to \infty$, we will have proved the required result. To do this we consider these four integrals.

Case I. Integral over Γ_1 or BJ
Along Γ_1 we have, since $s = Re^{i\theta}$, $\theta_0 \le \theta \le \pi/2$

$$I_1 = \int_{\Gamma_1} e^{st} f(s) ds = \int_{\theta_0}^{\pi/2} e^{Re^{i\theta}t} f\left(Re^{i\theta}\right) iRe^{i\theta} d\theta$$

Then

$$\begin{aligned} |I_1| &\le \int_{\theta_0}^{\pi/2} \left|e^{(R\cos\theta)t}\right| \left|e^{i(R\sin\theta)t}\right| \left|f\left(Re^{i\theta}\right)\right| \left|iRe^{i\theta}\right| d\theta \\ &\le \int_{\theta_0}^{\pi/2} e^{(R\cos\theta)t} \left|f\left(Re^{i\theta}\right)\right| R d\theta \\ &\le \frac{M}{R^{k-1}} \int_{\theta_0}^{\pi/2} e^{(R\cos\theta)t} d\theta = \frac{M}{R^{k-1}} \int_0^{\phi_0} e^{(R\sin\phi)t} d\phi \end{aligned}$$

Where we have used the given condition $|f(s)| \le M/R^k$ on Γ_1 and the transformation $\theta = \pi/2 - \phi$ where $\phi_0 = \pi/2 - \theta_0 = \arcsin(\gamma/R) = \sin^{-1}(\gamma/R)$.

Since $\sin\phi \le \sin\phi_0 \le \cos\theta_0 = \gamma/R$, this last integral is less than or equal to

$$\frac{M}{R^{k-1}} \int_0^{\phi_0} e^{\gamma t} d\phi = \frac{Me^{\gamma t}\phi_0}{R^{k-1}} = \frac{Me^{\gamma t}}{R^{k-1}} \sin^{-1}\frac{\gamma}{R}$$

But as $R \to \infty$, this last quantity approached zero as can be seen by noting, for example, that $\sin^{-1}(\gamma/R) \approx \gamma/R$ for large R. Thus, $\lim_{R\to\infty} I_1 = 0$.

Case II. Integral over Γ_2 or JPK
Along Γ_2 we have, since $s = Re^{i\theta}$, $\pi/2 \le \theta \le \pi$

$$I_2 = \int_{\Gamma_2} e^{st} f(s) ds = \int_{\pi/2}^{\pi} e^{Re^{i\theta}t} f\left(Re^{i\theta}\right) iRe^{i\theta} d\theta$$

Then, as in Case 1, we have

(continued)

$$|I_2| \le \frac{M}{R^{k-1}} \int_{\pi/2}^{\pi} e^{(R\cos\theta)t} d\theta = \frac{M}{R^{k-1}} \int_0^{\pi/2} e^{-(R\sin\phi)t} d\phi$$

Upon letting $\theta = \pi/2 + \phi$

Now $\sin\phi \ge 2\phi/\pi$ for $0 \le \varphi \le \pi/2$ so that the last integral is less than or equal to

$$\frac{M}{R^{k-1}} \int_0^{\pi/2} e^{-2r\phi t/\pi} d\phi = \frac{\pi M}{2tR^k} \left(1 - e^{Rt}\right)$$

which approaches zero as $R \to \infty$. Thus, $\lim_{R\to\infty} I_2 = 0$

Case III. Integral over Γ_3 or KQL
This case can be treated in a manner similar to Case II in above.

Case IV. Integral over Γ_4 or QL
This case can be treated in a manner similar to Case I in above.

Problem 2: Find the Inverse Laplace transform of the following function with branch point by use of complex inversion formula.

$$£^{-1}\left\{\frac{e^{-a\sqrt{s}}}{s}\right\} = ?$$

Solution: By the complex inversion formula and use of Fig. (b) below, the required inverse Laplace transform is given by:

$$F(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{st-a\sqrt{s}}}{s} ds \tag{1}$$

Since $s = 0$ is a branch point of the integrand, we consider

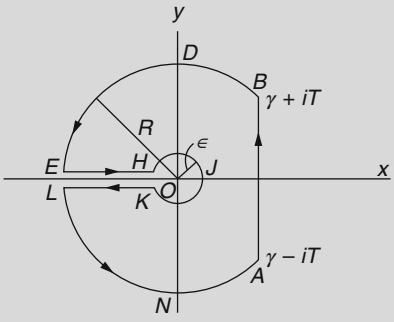


Fig. b Branch Contour

(continued)

$$\frac{1}{2\pi i}\oint_C \frac{e^{st-a\sqrt{s}}}{s}ds = \frac{1}{2\pi i}\int_{AB}\frac{e^{st-a\sqrt{s}}}{s}ds + \frac{1}{2\pi i}\int_{BDE}\frac{e^{st-a\sqrt{s}}}{s}ds$$
$$+\frac{1}{2\pi i}\int_{EH}\frac{e^{st-a\sqrt{s}}}{s}ds + \frac{1}{2\pi i}\int_{HJK}\frac{e^{st-a\sqrt{s}}}{s}ds$$
$$+\frac{1}{2\pi i}\int_{KL}\frac{e^{st-a\sqrt{s}}}{s}ds + \frac{1}{2\pi i}\int_{LNA}\frac{e^{st-a\sqrt{s}}}{s}ds$$

where C is the contour of Fig. (b) consisting of the line AB ($s=\gamma$), the arcs BDE and LNA of a circle of radius R and the center at origin O, and the arc HJK of a circle of radius ε with center at O.

Since the only singularity $s=0$ of the integrand is not inside C, the integral on the left is zero by Cauchy's theorem. Also, the integrand satisfies the condition of Problem 1 in above, so that on taking the limit as $R\to\infty$ the integrals along BDE and LNA approach zero. It follows that

$$F(t) = \lim_{\substack{R\to\infty\\ \varepsilon\to 0}} \frac{1}{2\pi i}\int_{AB}\frac{e^{st-a\sqrt{s}}}{s}ds = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{e^{st-a\sqrt{s}}}{s}ds$$
$$= \lim_{\substack{R\to\infty\\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{\int_{EH}\frac{e^{st-a\sqrt{s}}}{s}ds + \int_{HJK}\frac{e^{st-a\sqrt{s}}}{s}ds + \int_{KL}\frac{e^{st-a\sqrt{s}}}{s}ds + \right\} \tag{2}$$

Along EH, $s=xe^{\pi i}$, $\sqrt{s}=\sqrt{x}e^{\pi i/2}=i\sqrt{x}$ and as s goes from $-R$ to $-\varepsilon$, x goes from R to ε. Hence, we have

$$\int_{EH}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{-R}^{-\varepsilon}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{R}^{\varepsilon}\frac{e^{-xt-ai\sqrt{x}}}{x}dx$$

Similarly, along KL, $s=xe^{-\pi i}$, $\sqrt{s}=\sqrt{x}e^{-\pi i/2}=-i\sqrt{x}$ and as s goes from $-\varepsilon$ to $-R$, x goes from ε to R. Then, we have

$$\int_{KL}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{-\varepsilon}^{-R}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{\varepsilon}^{R}\frac{e^{-xt+ai\sqrt{x}}}{x}dx$$

Along HJK, $s=\varepsilon e^{i\theta}$ and we have

(continued)

$$\int_{\text{HJK}} \frac{e^{st-a\sqrt{s}}}{s} ds = \int_{\pi}^{-\pi} \frac{e^{\varepsilon e^{i\theta} t - a\sqrt{\varepsilon} e^{i\theta/2}}}{\varepsilon e^{i\theta}} i\varepsilon e^{i\theta} d\theta$$
$$= i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta} t - a\sqrt{\varepsilon} e^{i\theta/2}} d\theta$$

Thus, (2) becomes

$$F(t) = -\lim_{\substack{R\to\infty \\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{ \int_R^{\varepsilon} \frac{e^{-xt-ai\sqrt{x}}}{x} dx + \int_{\varepsilon}^{R} \frac{e^{-xt+ai\sqrt{x}}}{x} dx + i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta} t - a\sqrt{\varepsilon} e^{i\theta/2}} d\theta \right\}$$
$$= -\lim_{\substack{R\to\infty \\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{ \int_{\varepsilon}^{R} \frac{e^{xt}\left(e^{ai\sqrt{x}} - e^{-ai\sqrt{x}}\right)}{x} dx + i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta} t - a\sqrt{\varepsilon} e^{i\theta/2}} d\theta \right\}$$
$$= -\lim_{\substack{R\to\infty \\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{ 2i\int_{\varepsilon}^{R} \frac{e^{-xt}\sin a\sqrt{x}}{x} dx + i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta} t - a\sqrt{\varepsilon} e^{i\theta/2}} d\theta \right\}$$

Since the limit can be taken underneath the integral sign, we have

$$\lim_{\varepsilon\to 0} \int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta} t - a\sqrt{\varepsilon} e^{i\theta/2}} d\theta = \int_{\pi}^{-\pi} 1 \; d\theta = -2\pi$$

and so we find that:

$$F(t) = 1 - \frac{1}{\pi}\int_0^{\infty} \frac{e^{-xt}\sin a\sqrt{x}}{x} dx$$

and this result can be written as error function. See next Problem 3 below

$$F(t) = 1 - \text{erf}\left(a/2\sqrt{t}\right) = \text{erfc}\left(a/2\sqrt{t}\right)$$

Problem 3: Prove that

$$\frac{1}{\pi}\int_0^{\infty} \frac{e^{-xt}\sin a\sqrt{x}}{x} dx = \text{erf}\left(a/2\sqrt{t}\right)$$

Letting $x = u^2$, the required integral becomes

$$I = \frac{2}{\pi}\int_0^{\infty} \frac{e^{-u^2 t}\sin au}{u} du$$

(continued)

Then differentiating with respect to a and using the following

$$\frac{\partial I}{\partial a} = \frac{2}{\pi}\int_0^\infty e^{-u^2 t}\cos au\,du = \frac{2}{\pi}\left(\frac{\sqrt{\pi}}{2\sqrt{t}}e^{-a^2/4t}\right) = \frac{1}{\sqrt{\pi t}}e^{-a^2/4t}$$

Hence, using the fact that $I = 0$ when $a = 0$,

$$I = \int_0^a \frac{1}{\sqrt{\pi t}}e^{-p^2/4t}dp = \frac{2}{\sqrt{\pi}}\int_0^{a/2\sqrt{t}} e^{-u^2}du = \operatorname{erf}\left(a/2\sqrt{t}\right)$$

and the required result is established.

Now that we learned all above knowledge and using Fig. (c) here, let us now derive the formula and solution for the temperatures $T(z,t)$ in a semi-infinite solid $z \geq 0$, initially at temperature zero, when a constant flux of heat (F_0) is maintained at the boundary $z = 0$.

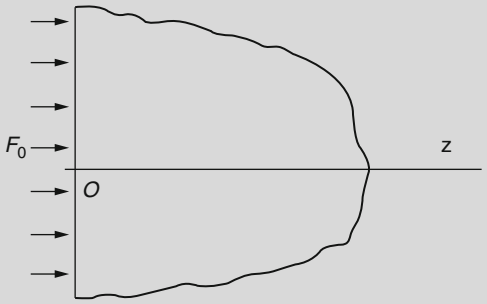


Fig. c

In this idealized case of a thick slab of material, we shall substitute for the thermal condition at the right-hand boundary, the condition that T tends to zero as z tends to infinity. The boundary value problem then becomes as:

1. $\dfrac{\partial^2 T(z,t)}{\partial z^2} - \dfrac{1}{\kappa}\dfrac{\partial T(z,t)}{\partial t} = 0 \qquad (z > 0\,,\, t > 0)$
2. $T(z,0) = 0 \qquad (z > 0)$
3. $-K\dfrac{\partial T(z,t)}{\partial z}\Big|_{z=0} = F_0 \qquad \lim_{z\to\infty} T(z,t) = 0 \qquad t > 0$

Let $u(z,s)$ be Laplace transform, with respect to time t, of the temperature function $T(z,t)$. Transforming the members of equations (1) and (3), we have the following problem in ordinary differential equations which $u(z,s)$ must satisfy, so we can write:

(continued)

$$su(z,s) - T(z,0) = \kappa \frac{d^2u(z,s)}{dz^2} \quad \text{or} \quad \frac{d^2u(z,s)}{dz^2} - \frac{s}{\kappa}u(z,s) = 0 \qquad (4)$$

where

$$u(0,s) = £\{T(o,t)\} = \frac{F_0}{s} \qquad (5)$$

and $u = u(z,s)$ is required to be bounded. Now the boundary condition (3) in the new Laplace plane becomes:

$$-K\frac{\partial T(z,s)}{\partial z} = \frac{F_0}{s} \quad \lim_{z\to\infty} T(z,s) = 0 \qquad (6)$$

Solving Eq. 4, we find that:

$$u(z,s) = c_1 e^{\sqrt{s/\kappa}z} + c_2 e^{-\sqrt{s/\kappa}z} \qquad (7)$$

Then we choose $c_1 = 0$ so that u is bounded as $z \to \infty$, and we have

$$u(z,s) = c_2 e^{-\sqrt{s/\kappa}z} \qquad (8)$$

From boundary condition of Eq. 6 and what is shown in Eq. 5 we have $c_2 = F_0\sqrt{\kappa}/Ks\sqrt{s}$, so that

$$u(z,s) = \frac{F_0\sqrt{\kappa}}{Ks\sqrt{s}} e^{-\sqrt{s/\kappa}z}$$

Hence, by Problems 2 and 3 in above, we know that:

$$£^{-1}\left\{\frac{1}{\sqrt{s}} e^{-z\sqrt{s/\kappa}}\right\} = \frac{1}{\sqrt{\pi t}} e^{-z^2/(4\kappa t)}$$

and view of the factor $1/s$ in our formula for $u(z,s)$ it follows

$$\begin{aligned} T(z,t) &= \frac{F_0}{K}\sqrt{\frac{\kappa}{\pi}}\int_{\tau=0}^{\tau=t} e^{-z^2/(4\kappa\tau)}\frac{d\tau}{\sqrt{\tau}} \\ &= \frac{F_0 z}{K\sqrt{\pi}}\int_{z/\left(2\sqrt{\kappa t}\right)}^{\infty} \frac{1}{\lambda^2} e^{-\lambda^2} d\lambda \end{aligned}$$

Where the second integral is obtained from the first by substitution $\lambda = z/\left(2\sqrt{\kappa t}\right)$. Upon integration the last integral by parts, we find that

(continued)

$$T(z,t) = \frac{F_0}{K\sqrt{\pi}}\left(2\sqrt{\kappa t}e^{-z^2/(4\kappa t)} - 2z\int_{z/\left(2\sqrt{\kappa t}\right.}^{\infty} e^{-\lambda^2}d\lambda\right)$$

Which is result of Eq. 7.160

The solution to this problem is given by:

$$T(z,t) = \frac{2F_0\sqrt{\kappa t}}{K}\,\text{ierf}\left[z/2\sqrt{\kappa t}\right] \tag{7.160}$$

or

$$T(z,t) = \frac{2F_0}{K}\left\{\sqrt{\frac{\kappa t}{\pi}}e^{-z/4\kappa t} - \frac{z}{2}\text{erf}\left[z/\left(2\sqrt{\kappa t}\right)\right]\right\}$$

The functions which appear here are error functions, and it is useful to summarize some of their properties and definitions. (See Carslaw and Jaeger Chapter II) [3].

The error function is:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-\ell^2}d\ell$$

$$\text{erf}(0) = 0 \qquad \text{erf}(\infty) = 1 \qquad \text{erf}(-x) = -\text{erf}(x)$$

The complementary error function is:

$$\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-\ell^2}d\ell \quad \text{erfc}(0) = 1$$

The integral of the complementary error function is:

$$\text{ierfc}(x) = \int_x^{\infty}\text{erfc}(\ell)d\ell$$

or

$$\text{ierfc}(x) = \frac{1}{\sqrt{\pi}}e^{-x^2} - x\text{erfc}(x)$$

$$\text{ierfc}(x) = \frac{1}{\sqrt{\pi}}e^{-x^2} - x + x\text{erf}(x)$$

Some derivations are useful:

$$\frac{\partial \mathrm{erf}(x)}{\partial x} = -\frac{\partial \mathrm{erfc}(x)}{\partial x} = \frac{2}{\sqrt{\pi}} e^{-x^2}$$

$$\frac{\partial^2 \mathrm{erf}(x)}{\partial x^2} = -\frac{\partial^2 \mathrm{erfc}(x)}{\partial x^2} = -\frac{4}{\sqrt{\pi}} x e^{-x^2}$$

$$\frac{\partial \mathrm{ierfc}(x)}{\partial x} = -\mathrm{erfc}(x)$$

$$\frac{\partial^2 \mathrm{ierfc}(x)}{\partial x^2} = \frac{2}{\sqrt{\pi}} e^{-x^2}$$

Now we can show that the boundary condition is satisfied. Using the first form of Eq. 7.160 yields that:

$$\frac{\partial T}{\partial z} = \frac{2F_0\sqrt{\kappa t}}{K}\left\{-\mathrm{erfc}\left[z/\left(2\sqrt{\kappa t}\right)\right]\right\}\frac{1}{2\sqrt{\kappa t}}$$

and since $\mathrm{erfc}(0) = 1$, therefore

$$\left.\frac{\partial T}{\partial z}\right|_{z=0} = -\frac{F_0}{K}$$

One can also show that Eq. 7.160 satisfies Eq. 7.157.

We can use Eq. 7.160 to show what the front surface temperature behavior is, under constant irradiation, by setting $z = 0$ so that:

$$T(0,t) = \frac{2F_0}{K}\sqrt{\frac{\kappa t}{\pi}} \tag{7.161}$$

As an illustration, let us calculate the time required to raise aluminum to its melting point for a power density of 5 k W/cm^2:

$$K = 2.3\,\mathrm{W/cm^2}$$

$$\kappa \approx 0.9\,\mathrm{cm^2/s}$$

$$T = T_{\mathrm{melt}} - T_{\mathrm{room}} \approx 600\,^{\circ}\mathrm{C}$$

$$F_0 = (1-R)I$$

$$(1-R) = 0.04$$

$$F_0 = 0.04 \times 5 \times 10^3 = 200\,\mathrm{W/cm^2}$$

Then $t = \left(\pi K^2 T^2\right)/\left(4F_0^2\kappa\right)$ yields $t \approx 42\,\text{s}$. In practice, it is very difficult to melt extremely thick slabs of aluminum with even a high-power laser, as these calculations suggest.

Equation 7.160. although derived for a very simple case, describes man, very important features of thermal response to lasers. First we shall define the diffusion length, which is useful in that it permits a wide variety of order-of-magnitude calculations to be made. The thermal diffusion length D is defined as:

$$D = 2\sqrt{\kappa t} \tag{7.162}$$

Strictly speaking, the thermal diffusion length is defined as the distance required for the temperature to drop to $1/e$ of its initial value and depends somewhat on the geometry and the boundary condition. For most purpose it is sufficient simply to take it as defined by Eq. 7.162. Looking at our Eq. 7.160, for example, we see that:

$$T(D,t) = \frac{2F_0}{K}\left\{\sqrt{\frac{\kappa t}{\pi e}}1 - \frac{D}{2}\text{erfc}(1)\right\}$$

Using table of erfc we get: $\text{erfc}(1) \approx 0.1573$, See Table 7.2 below as well

Table 7.2 Table of error function and its complementary

Value x	Erf (x)	Erfc (x)
0	0	1
0.05	0.0563720	0.9436280
0.1	0.1124629	0.8875371
0.15	0.1679960	0.8320040
0.2	0.2227026	0.7772974
0.25	0.2763264	0.7236736
0.3	0.3286268	0.6713732
0.35	0.3793821	0.6206179
0.4	0.4283924	0.5716076
0.45	0.4754817	0.5245183
0.5	0.5204999	0.4795001
0.55	0.5633234	0.4366766
0.6	0.6038561	0.3961439
0.65	0.6420293	0.3579707
0.7	0.6778012	0.3221988
0.75	0.7111556	0.2888444
0.8	0.7421010	0.2578990
0.85	0.7706681	0.2293319
0.9	0.7969082	0.2030918
0.95	0.8208908	0.1791092
1	0.8427008	0.1572992

(continued)

Table 7.2 (continued)

Value x	Erf (x)	Erfc (x)
1.1	0.8802051	0.1197949
1.2	0.9103140	0.0896860
1.3	0.9340079	0.0659921
1.4	0.9522851	0.0477149
1.5	0.9661051	0.0338949
1.6	0.9763484	0.0236516
1.7	0.9837905	0.0162095
1.8	0.9890905	0.0109095
1.9	0.9927904	0.0072096
2	0.9953223	0.0046777
2.1	0.9970205	0.0029795
2.2	0.9981372	0.0018628
2.3	0.9988568	0.0011432
2.4	0.9993115	0.0006885
2.5	0.9995930	0.0004070
2.6	0.9997640	0.0002360
2.7	0.9998657	0.0001343
2.8	0.9999250	0.0000750
2.9	0.9999589	0.0000411
3	0.9999779	0.0000221
3.1	0.9999884	0.0000116
3.2	0.9999940	0.0000060
3.3	0.9999969	0.0000031
3.4	0.9999985	0.0000015
3.5	0.9999993	0.0000007

$$T(D,t) = \frac{2F_0}{K}\left\{\sqrt{\frac{\kappa t 1}{\pi e}} - \frac{D}{2}(0.1573)\right\}$$

From Eq. 7.161, we have:

$$T(D,t) = T(0,t)\left\{\frac{1}{e} - 0.1573\sqrt{\pi}\right\}$$

Thus

$$T(D,t) = 0.09T(0,t) \tag{7.163}$$

In this case, whereas $(1/e)T(0,t) \approx 0.37T(0,t)$. Referring to our example of irradiating aluminum for 42 s to reach the melting point, we note that the diffusion length at that time is given by:

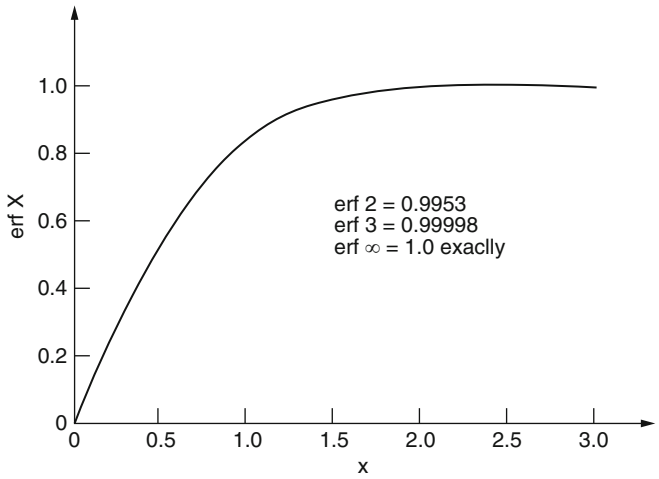


Fig. 7.29 The error function plot

$$D = 2\sqrt{0.9 \times 42} \approx 12 \text{ cm}$$

and by Eq. 7.163, and the value of $T = 600\,^{\circ}\text{C}$, the temperature at distance D into the material is 0.09×600, or about 54 °C above ambient temperature given by above values.

Now let us illustrate the solution Eq. 7.160 graphically as demonstrated in Fig. 7.29. For convenience rewrite the equation by introducing $D = 2\sqrt{\kappa t}$ and by reducing it to the error function erf, so that:

$$T(z,t) = \frac{2F_0}{K}\left[\frac{D}{2\sqrt{\pi}}e^{-z^2/D^2} - \frac{z}{2} + \frac{z}{2}\text{erf}(z/D)\right]$$

Now let $\eta = z/D$, and

$$T(z,t) = \frac{F_0 D}{K}\left[\frac{e^{-\eta^2}}{2\sqrt{\pi}} - \eta + \eta\text{erf}(\eta)\right] \tag{7.164a}$$

or equivalently in terms of ierfc, we can rewrite the above equation as follow:

$$T(z,t) = \frac{F_0 D}{K}\text{ierfc}(\eta) \tag{7.164b}$$

Finally, we define a dimensionless temperature $\theta = TK/(F_0 D)$, that [48]:

$$\theta = \text{ierfc}(\eta) \tag{7.164c}$$

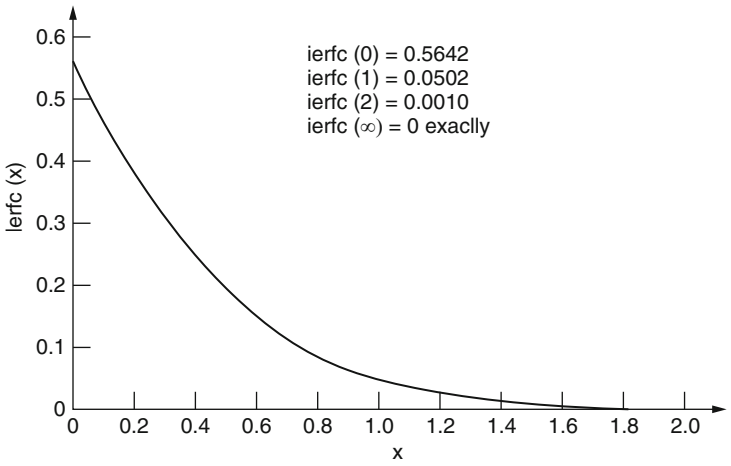


Fig. 7.30 The integral of the complementary error function

Thus, the plot of the integral of the complementary error function here is the graph of the solution to the problem of constant heat flux on the surface of a semi-infinite solid.

Now, although the graph of Eq. 7.164c (see Fig. 7.30) represents very succinctly the solution to our problem, it does not really snow how the temperature varies as a function of position and time. For this purpose it is useful to look at the temperature profiles for various times and see how the profile changes with time. These curves can be generated quickly from $\theta = \text{ierfc}(\eta)$ by recalling the definitions of θ and η, and writing them in the following form:

$$T = \left(\frac{2F_0\sqrt{\kappa}}{K}\right)\sqrt{t}\theta \tag{7.165a}$$

$$z = 2\sqrt{\kappa}\sqrt{t}\eta \tag{7.165b}$$

Thus, at a given time the $\theta = \text{ierfc}(\eta)$ curve scales according to Eq. 7.166; the basic shape of the curve is unchanged, but it is stretched one way or the other depending on the parameters, and this stretching progresses in time as $\sqrt{t}$. The case of aluminum is shown in Fig. 7.31.

We can also look at the variation of temperature with time at a fixed position. The variation of the surface $(z = 0)$ is simply $T(0, t) \approx \sqrt{t}$ as Eq. 7.161 shows. Wherever $z/\left(2\sqrt{\kappa t}\right)$ is very small, the temperature variation will approach $\sqrt{t}$. Thus, at any position $T \approx \sqrt{t}$ at sufficiently large t. The temperature vs. time profiles at fixed position for times such that $z/\left(2\sqrt{\kappa t}\right)$ is not small can be calculated, of course, from Eq. 7.160. Some results for aluminum, with the parameters used above, are shown in Fig. 7.32. Notice that at $z = 10$ cm the temperature profile is far from the

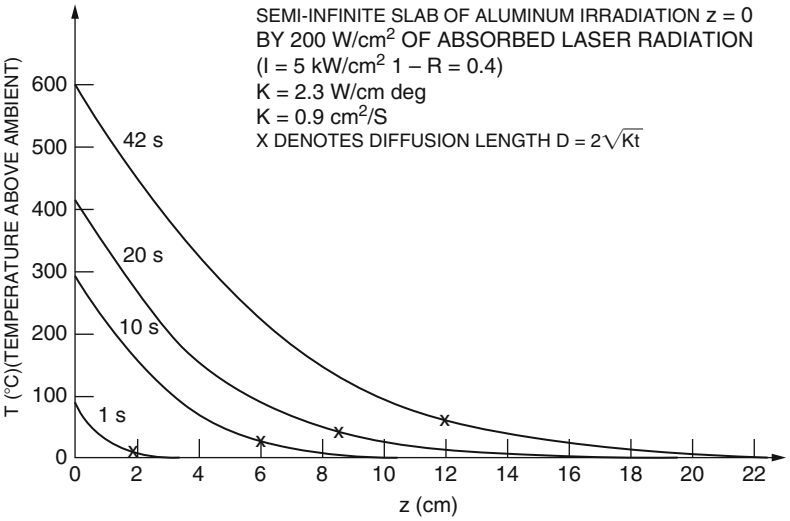

SEMI-INFINITE SLAB OF ALUMINUM IRRADIATION z = 0
BY 200 W/cm^2 OF ABSORBED LASER RADIATION
(I = 5 kW/cm^2 1 – R = 0.4)
K = 2.3 W/cm deg
K = 0.9 cm^2/S
X DENOTES DIFFUSION LENGTH D = $2\sqrt{Kt}$
T (°C)(TEMPERATURE ABOVE AMBIENT)
42 s
20 s
10 s
1 s
z (cm)

Fig. 7.31 Laser-induced temperature rise in aluminum as a function of depth

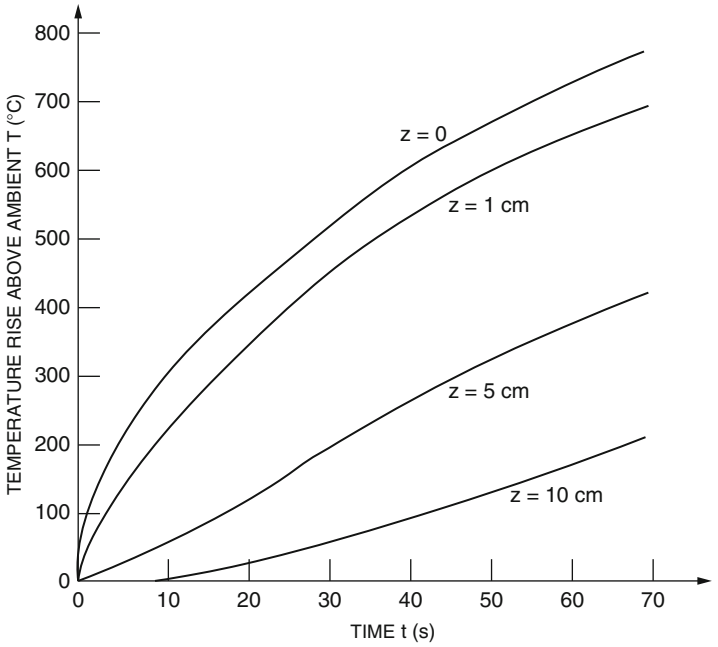


Fig. 7.32 Laser-induced temperature rise on front surface of aluminum as a function of time

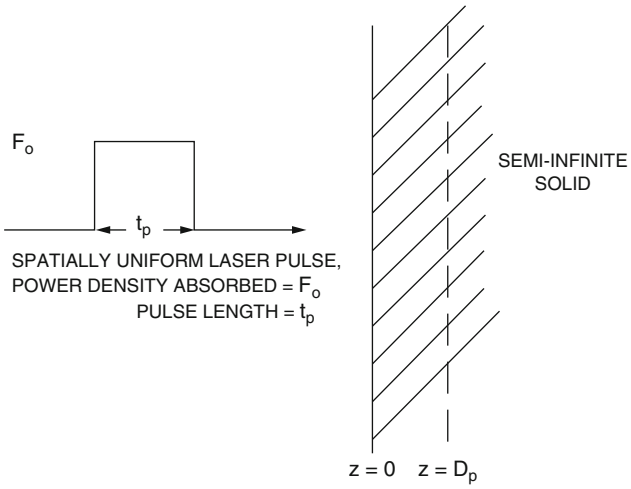


Fig. 7.33 Irradiation of a semi-infinite solid by a pulse

"long-time," or $\sqrt{t}$, behavior even at 40 or 50 s, whereas the surface has already begun to melt.

We now turn to some order-of-magnitude arguments. One such argument can be used to estimate the power-pulse length combination which might be expected to yield surface vaporization.

Consider a laser pulse that has the simple time behavior shown in Fig. 7.33 and uniformly irradiates the surface of the material. The pulse length is t_p and the intensity is such that, combined with the reflectance, the absorbed power density is F_0. Again we assume that the optical energy is absorbed in a very thin layer at the surface. Let D_p be the diffusion length associated with the time t_p.

The question is whether a significant amount of surface vaporization will occur before the pulse ends. One approach would be to use Eq. 7.161 to calculate the surface temperature at the time t_p and compare it to this vaporization temperature. However, this would ignore the influence of the latent heats of melting and of vaporization, which have an important influence. We shall discuss thermal flow with phase changes later. For the present purpose we can include them by considering the energy required to melt and vaporize a portion of the material. The key is to estimate what thickness of the material is involved, and in this order-of-magnitude argument we simply use the thermal diffusion length for this thickness. Thus, we set the criterion for vaporization as

$$\frac{F_0 t_\mathrm{p}}{D_\mathrm{p}} \geq \rho[c_\mathrm{s}(T_\mathrm{m} - T_0) + L_\mathrm{m} + c_\ell(T_\mathrm{b} - T_\mathrm{m}) + L_\upsilon]$$

where ρ is density of the material, c_s and c_ℓ are the specific heats of the solid and liquid, respectively, Tm is the melting point, T_b is the boiling point, and L_m and L_υ

are the heats of melting and vaporization. Notice we are ignoring differences between the solid and liquid for density and conductivity, as is appropriate in this crude argument. If numerical values are checked, L_v, dominates the expression on the right side of the inequality. For example, for aluminum $L_v = 10,875\,\mathrm{J/g}$, whereas all the other terms contribute a total of 3.046 J/g. Since the argument is crude, then, one usually takes:

$$\frac{F_0 t_\mathrm{p}}{D_\mathrm{p}} \geq \rho L_v$$

as the criterion for vaporization by a pulse. Since $D_\mathrm{p} = 2\sqrt{\kappa t_\mathrm{p}}$, we have

$$F_0 \geq \text{aprrox } \frac{2\sqrt{\kappa} L_v \rho}{\sqrt{t_\mathrm{p}}} \tag{7.166}$$

Some calculations based on Eqs. 7.165a and 7.165b are shown in Fig. 7.34. Most metals fall in the band indicated. For a given pulse time, at power densities greater than the band indicates, vaporization effects would be expected to be important. Some useful thermal constants are included in Table 7.3.

In deriving Eq. 7.166, we have been seeking the power density required, at a given pulse length, for a thermal layer to be vaporized. The same expression, of course, tells us the pulse time at which vaporization becomes important for a fixed power density. Rewriting Eq. 7.166 gives for this time

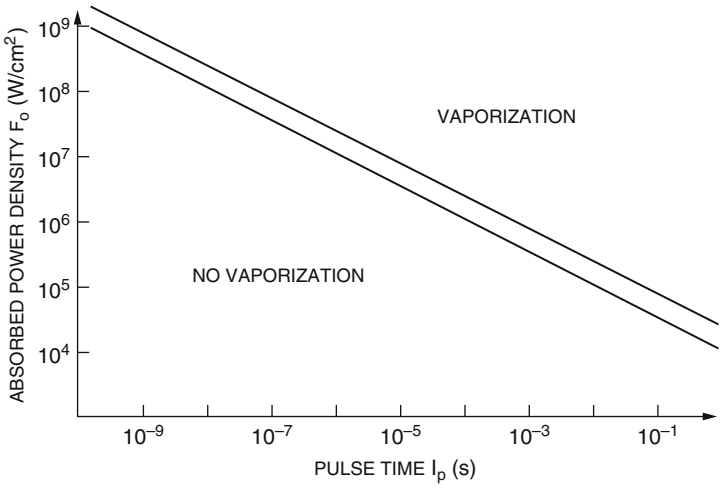


Fig. 7.34 Power density pulse time criterion for vaporization

Table 7.3 Thermochemical data for metals[a]

Metal or alloy	Diffusivity[b] (cm^2/s)	Conductivity[b]		Specific heat[b,c]		Density ((g/cm^3))	Solidus temperature (°C)	Liquidus temperature (°C)	Heat of melting		Vaporization temperature (°C)	Heat of vaporization	
		(W/cm K)	(cal/s cm K)	(J/g K)	(cal/g K)				(J/g)	(cal/g)		(J/g)	(cal/g)
Al	0.85	2.40	0.57	1.05	0.25	2.7		660	400	95.6	2520	10,875	2600
2024 Al	0.60	1.732	0.414	1.0	0.24	2.77	502	638					
Ti	0.0636	0.216	0.0516	0.753	0.18	4.51		1670	324	77.3	3289	8790	2100
Ti (6A1, 4 V)	0.051	0.19	0.0455	0.837	0.20	4.47	1537	1649					
Fe						7.86		1536	247	59.1	2862	6260	1496
304 Stainless Steel	0.0523	0.259	0.062	0.628	0.15	8.0	1399	1454					

[a]Data compiled by R. L. Stegman of the Naval Research Laboratory. Washington, DC
[b]These values represent an average over the temperature between room temperature and the melting point
[c]Of the solid

$$t_p > \text{aprrox } \frac{K^2 T_{vap}^2 \pi}{4F_0^2 \kappa} \tag{7.167}$$

Let us compare this to the time required for surface vaporization to begin. We do this by using our solution for heat flow in the semi-infinite solid for the surface (Eq. 7.161) and solving for the time at which the front surface reaches the vaporization temperature:

$$T_{vap} = \frac{2F_0}{K}\sqrt{\frac{\kappa t_{vap}}{\pi}}$$

or

$$t_{vap} = \frac{K^2 T_{vap}^2 \pi}{4F_0^2 \kappa}$$

Thus, at $t_p = t_{vap}$ this calculation would predict that vaporization at the surface begins. For example, at $F_0 = 10^6\,\text{W/cm}^2$, vaporization begins at $t_p \approx 10^{-5}$ to $10^{-6}\,\text{s}$, depending on the metal. On the other hand, for a thermal layer to be evaporated requires, according to Eq. 7.167, $t_p \approx 10^{-3}\,\text{s}$. It turns out that both estimates are useful. In a later section we shall discuss features of a more correct treatment, which accounts for both the heat of melting and the heat of vaporization in the dynamic situation of propagating solid–liquid and liquid–vapor interfaces.

7.9.1.2 No Phase Change: Slab of Finite Thickness

Let us turn now to a treatment of another geometry which can be useful in practical cases, namely irradiation of one surface of a sheet or slab of finite thickness. Let the slab be taken as infinite in extent in the x and y direction, and let the laser irradiation be uniform over the entire surface $z = 0$. Thus, we again have a one-dimensional situation, as shown in Fig. 7.35. The thickness of the sheet is taken as ℓ, the absorbed power density as a function of time is $F(t)$, and we again assume that the radiation is absorbed in a very narrow layer at the front surface. The equation we wish to solve is, then:

$$\frac{\partial^2 T(z,t)}{\partial z^2} - \frac{1}{\kappa}\frac{\partial T(z,t)}{\partial t} = 0$$

with boundary conditions

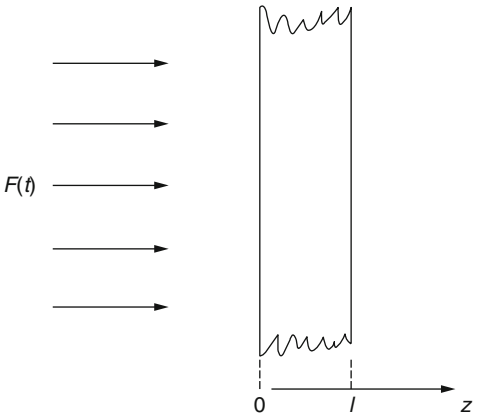


Fig. 7.35 Irradiation of slab of finite thickness

$$-K\frac{\partial T(z,t)}{\partial z}\bigg|_{z=0} = F(t)$$

$$-K\frac{\partial T(z,t)}{\partial z}\bigg|_{z=\ell} = 0$$

The second boundary condition states that—he rear surface is insulated. We shall look at the consequences of this assumption a little later.

As we showed for the semi-infinite slab, the solutions turn out to be elaborate. Turning to the special case of $F(t) = F_0$, a constant, the solution is as follows and easily can be obtained very similar to the methods in case I by using Laplace Transform or separation of variables methods by assuming $T(z,t) = Z(z)U(t)$ and given boundary conditions to obtain a general solution. In case Laplace Transformation approach the Ordinary Differential Equation (ODE) in Laplace time and space plane is given as:

$$su(z,s) - T(x,0) = \kappa\frac{d^2u(z,s)}{dz^2} \quad \text{or} \quad \frac{d^2u(z,s)}{dz^2} - \frac{su(z,s)}{\kappa} = -\frac{F_0}{\kappa} \tag{7.168}$$

Taking Laplace Transform over the boundary conditions provides the new form of them in Laplace domain as:

$$£\left\{-K\frac{\partial T(z,t)}{\partial z}\bigg|_{z=0}\right\} = \frac{\partial u(z,s)}{\partial z}\bigg|_{z=0} = \frac{F_0}{Ks}$$

$$£\left\{-K\frac{\partial T(z,t)}{\partial z}\bigg|_{z=\ell}\right\} = \frac{\partial u(z,s)}{\partial z}\bigg|_{z=\ell} = 0$$

The general solution of Eq. 7.168 is written as:

$$u(z,s) = c_1 \cosh\sqrt{s/\kappa}z + c_2 \sinh\sqrt{s/\kappa}z$$

Applying the new sets of boundary conditions to solve for c_1 and c_2 and taking reverse Laplace Transform to transfer from coordinates of (z,s) to (z,t) as well as dealing with finding the residues at given s ,then calculating the sum of residues and some mathematical manipulations we will end up with the following solution (Eq. 7.169). Readers should go through this exercise on own.

$$T(z,t) = \frac{F_0\kappa}{K\ell}t + \frac{F_0\ell}{K}\left\{\frac{3(\ell-z)^2-\ell^2}{6\ell^2} - \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}e^{-\kappa n^2\pi^2 t/\ell^2}\cos\left[\frac{n\pi(\ell-z)}{\ell}\right]\right\} \tag{7.169}$$

We can check that this satisfies the boundary conditions:

$$\frac{\partial T(z,t)}{\partial z} = \frac{F_0\ell}{K}\left\{\frac{-(\ell-z)}{\ell^2} - \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}e^{-\kappa n^2\pi^2 t/\ell^2}\frac{n\pi}{\ell}\sin\left[\frac{n\pi(\ell-z)}{\ell}\right]\right\}$$

Now $\sin(n\pi) = 0$ and $\sin(0) = 0$, so the $\sum$ term vanishes at both $z = 0$ and $z = \ell$, and

$$\left.\frac{\partial T(z,t)}{\partial z}\right|_{z=0} = -\frac{F_0}{K}$$

$$\left.\frac{\partial T(z,t)}{\partial z}\right|_{z=\ell} = 0$$

Similarly, the thermal diffusion equation is satisfied, as the reader can verify.

Let us look briefly at this solution. It consists of a linear term in t, together with a "correcting term," which can he plotted as shown in Fig. 7.36. In other words, what is plotted is the term

$$D = \frac{3(\ell-z)^2-\ell^2}{6\ell^2} - \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}e^{-\kappa n^2\pi^2 t/\ell^2}\cos\left[\frac{n\pi(\ell-z)}{\ell}\right]$$

Let us examine some special cases. At $z = 0$,for example, we have:

$$T(0,t) = \frac{F_0\kappa}{K\ell}t + \frac{F_0\ell}{K}\left[\frac{1}{3} - \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}e^{-\kappa n^2\pi^2 t/\ell^2}\right] \tag{7.170}$$

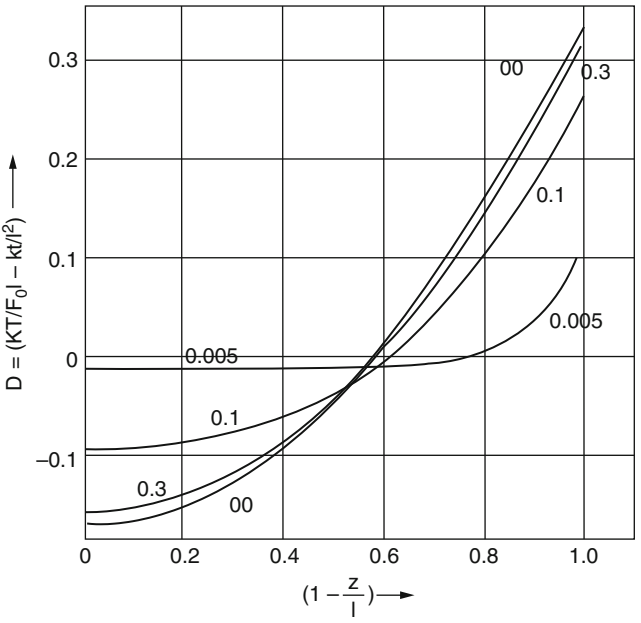


Fig. 7.36 D as a function of $1 - z/\ell$ (reference [3], Page 113)

We can rewrite this as

$$T(0,t) = \frac{F_0\kappa}{K\ell}t + \frac{F_0\ell}{K}D_{z=0}$$

Now D at a fixed value of z is a function of $\kappa t/\ell^2$. If $\eta = \kappa t/\ell^2$, we can write

$$T(0,t) = \frac{F_0\ell}{K}[\eta + D_{z=0}(\eta)] \tag{7.171}$$

Figure 7.37 shows how D, depends on , η, for $z = 0$, and was taken from the previous graph of D, vs. $(1 - z/\ell)$ as per Fig. 7.36. Note that at small η, i.e., at $\kappa t \ll \ell^2$, $D \approx 0$, so that the front surface initially heats up linearly with time, as:

$$T(0,t) = \frac{F_0\kappa}{K\ell}t \tag{7.172}$$

For large t values, or $\kappa t \gg \ell^2$, $D_{z=0}$ approaches a limiting value of about 0.33. Thus, at long times, we have:

$$T(0,t) = \frac{F_0\ell}{K}\left[\frac{\kappa t}{\ell^2} + 0.33\right] \quad \kappa t \gg \ell^2 \tag{7.173}$$

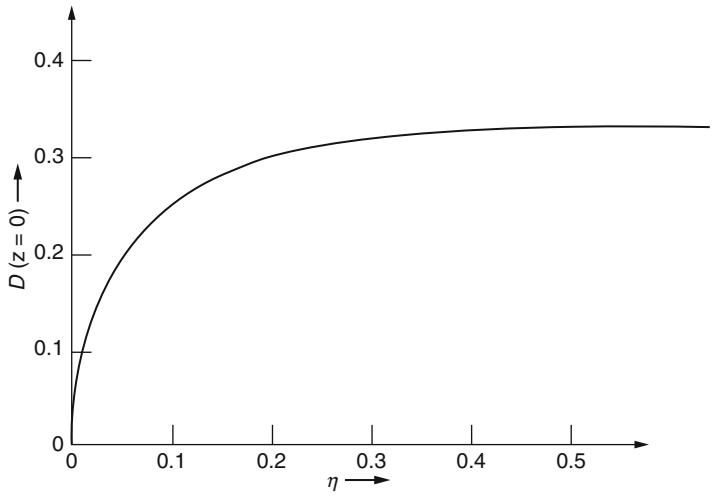


Fig. 7.37 D at z=0 as a function of η

Here again we see linear behavior, but this time there is an additive constant. If we have a very thick slab, we should get the same result as our previous solution for the time to reach 600 °C on the front surface of aluminum with an absorbed power density of 200 W/cm^2. It turns out that the limiting form of Eq. 7.172 is not correct because it ignores the behavior of $D_{z=0}(\eta)$ at small η. It. is necessary to use the full expression. Thus

$$\eta + D_{z=0}(\eta) = \frac{KT}{F_0 \ell}$$

Assume that $\ell = 100$ cm, since we know from our infinite-slab solution that the diffusion distance is 12 cm at $T = 600$ °C on the front. Thus

$$\eta + D_{z=0}(\eta) = \frac{2.3 \times 600}{200 \times 100} = 0.069$$

Reading very rough from the graph of $D_{z=0}$ vs. η gives

$$\eta \approx 0.004 \quad \text{at} \quad D \approx 0.065$$

Thus, our solution is:

$$\eta \approx 0.004 = \frac{\kappa t}{\ell^2}$$

which gives, since $\kappa \approx 0.9\,\mathrm{cm}^2/\mathrm{s}$, a time of about 44 s, in reasonable agreement with the semi-infinite-slab solution.

Now let us turn to a consideration of the rear surface temperature. For this case $z = \ell$, so Eq. 7.169 reduces to:

$$T(z,t) = \frac{F_0 t\kappa}{K\ell} + \frac{F_0 \ell}{K}\left\{-\frac{1}{6} - \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}e^{-\kappa n^2\pi^2 t/\ell^2}\right\}$$

or, introducing η as before, we have:

$$T(\ell,t) = \frac{F_0\ell}{K}[\eta + D_{z=\ell}(\eta)]$$

Comments could be made here for the rear surface temperature, and they would be similar to those we made for the front surface temperature. It is interesting to compare the front surface temperature to the back surface temperature. This has a simple form for thin sheets, where $\kappa t/\ell^2 \gg 1$. By referring to the graph Fig. 7.36 of D, vs. $(1 - z/\ell)$, one can read off values for $D_{z=\ell}(\infty)$ and $D_{z=0}(\infty)$, and thus:

$$T(0,t) - T(\ell,t) \approx 0.5\frac{F_0\ell}{K}$$

for

$$\frac{\kappa t}{\ell^2} \gg 1$$

Notice in Fig. 7.36 that the limiting values are approached rapidly; they are nearly realized by the time $\kappa t/\ell^2 \gg 1$. As a numerical illustration, if we have 0.3 cm-thick aluminum,

$$T(0,t) - T(\ell,t) \approx 13^\circ\mathrm{C}$$

For the same numbers we used above. This situation, with the two surfaces heating at the same rate but separated by 13 °C, would start at a time of the order of $t \approx \ell^2/\kappa \approx 0.1$ s. At this time the front surface temperature is about 35 °C.

Let us turn now to a different sort of heat input. So far we have been discussing continuous irradiation. Another simple case, which is a reasonable approximation under certain conditions, is that of a laser pulse which is short enough to be treated as a delta function. Take again a slab of thickness ℓ, and assume that the energy is deposited in a very thin layer near the surface. F refers, as before, to the fraction absorbed by the material. The laser power density, that is, must be multiplied by the optical absorptance. In this case we solve the thermal diffusion equation subject to the boundary condition that:

$$-K\frac{\partial T(z,t)}{\partial z}\bigg|_{z=0} = -K\frac{\partial T(z,t)}{\partial z}\bigg|_{z=\ell} = 0$$

with the stipulation that there is an instantaneous release of E0 units of energy per unit area in the plane $z = 0$ at time zero. This type of problem is discussed in Carslaw and Jaeger [3] and is most easily solved by Laplace transform methods. For our present purpose we quote the solution

$$T(z,t) = \frac{F_0\kappa}{K\ell}\left\{1 + 2\sum_{n=1}^{\infty}\cos\left(\frac{n\pi z}{\ell}\right)e^{-\kappa n^2\pi^2 t/\ell^2}\right\} \tag{7.174}$$

In Eq. 7.174 we have introduced E_0, the energy per unit area in the pulse. Thus,

$$E_0 = \int_0^{\infty} F(t)dt$$

For the case under consideration, $F(t)$ is considered to be a delta function.

Equation 7.174 is the basis for a scheme used quite frequently for the measurement of thermal parameters [49]. This scheme consists of using a thin sheet of the material to be studied and irradiating uniformly one surface with a very short laser pulse while monitoring the temperature rise induced on the back surface. If one knows E_0, and if the assumptions of no heat loss are valid, the experiment can yield values of both specific heat and thermal conductivity. One adjusts the pulse energy, and hence E_0, so that the induced temperature rise is small. In this way the values of specific heat and thermal conductivity are representative of essentially the ambient temperature of the material.

To see how this is applied, rewrite Eq. 7.174 for the back surface, $z = \ell$ as:

$$T(z,t) = \frac{F_0\kappa}{K\ell}\left\{1 + 2\sum_{n=1}^{\infty}(-1)^n e^{-\kappa n^2\pi^2 t/\ell^2}\right\} \tag{7.175}$$

If we introduce a characteristic time $t_c = \ell^2/\kappa\pi^2$, Eq. 7.175 looks approximately like the curve shown in Fig. 7.38. Here we have also introduced a characteristic temperature $t_c = E_0\kappa/K\ell^2$ and plotted T/T_c, vs. t/t_c, or

$$\frac{T}{T_c} = 1 + 2\sum_{n=1}^{\infty}(-1)^n e^{-n^2 t/t_c} \tag{7.176}$$

Essentially the experiment consists of monitoring the temperature as a function of time and fitting it to Eq. 7.177. This can be done quite readily. First, the long-term temperature rise T_∞ yields the specific heat because:

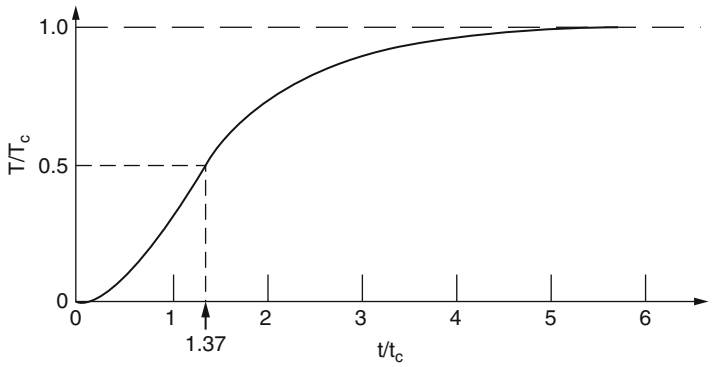


Fig. 7.38 Normalized back surface temperature response to a delta function heat pulse

$$\frac{T_\infty}{T_c} = 1$$

and, on substituting for T_c

$$T_\infty = \frac{E_0 \kappa}{K\ell}$$

or, since $\kappa = K/\rho c_p$

$$c_p = \frac{E_0}{T_\infty \rho \ell} \tag{7.177}$$

This technique of measuring specific heat is, of course, not unique, to pulsed lasers. It is sometimes referred to as the slat, calorimeter. The accuracy of the method depends on knowing E_0, which is frequently difficult to ascertain with laser radiation. In some application Eq. 7.172 is used to calculate E_0, the energy actually absorbed from the laser pulse, by using materials o. known specific heat.

The pulsed laser measurement technique is especially suited to determining thermal diffusivity. The magnitude of the back surface temperature rise depends, as we saw, on the energy which is coupled into the material, and this may be difficult to know with any precision. However the time dependence of the hack surface temperature is independent of the energy input and is controlled only by the diffusivity κ. A simple way to derive κ from a temperature–time profile can lie seen from Fig. 7.38. One measure the time required for the measured temperature response to reach some fraction, say one-half, of its limiting value. Let us call this time Il. It can lie shown numerically [49] from Eq. 7.176 that:

$$\frac{T}{T_c} = \frac{1}{2} = 1 + 2\sum_{n=1}^{\infty} (-1)^n e^{-n^2 t_{1/2}/t_c}$$

is satisfied when

$$\frac{t_{1/2}}{t_c} = 1.37$$

Using $t_c = \ell^2/(\kappa\pi^2)$ yields

$$\kappa = \frac{1.37}{\pi^2} \frac{\ell^2}{t_{1/2}} \tag{7.178}$$

Thus, a measurement of $t_{1/2}$, together with the thickness of the specimen, immediately gives thermal diffusivity. If one knows E_0 this experiment gives values of both the, specific heat and the diffusivity, and hence, if one knows the density ρ the experiment gives the thermal conductivity.

This technique has been applied often, at rather high temperatures. usually in the 1000 °C range and above. At these temperatures steady-state methods of measuring the thermal conductivity are difficult to apply because radiation losses are so large. In the laser flash technique the radiation loss goes like $T_\infty^4 - T_0^4$ where T_0 is the starting or ambient temperature, established by, say, a furnace, and T_∞ is defined above. This radiation loss can be made quite small by adjusting E_0 so that T_∞ is only a few degrees larger than T_0. Since 0the precise value of E_0 is difficult to establish, these experiments typically measure only thermal diffusivity, not the conductivity.

A final remark on the criterion for the applicability of Eq. 7.177 for slab heating concerns the limits on the laser pulse duration. No laser pulse is, of course, a true δ function. Our solution would be expected to be correct for laser pulse times which are short compared to the time it takes the back surface to respond. The response times are of the order of t_c, so we have the criterion as:

$$t_p \ll t_c$$

or

$$t_p \ll \ell^2/(\kappa\pi^2)$$

Calculations that include explicitly the time dependence of the laser pulse [50] indicate that our δ function solution is in error by less than about 2 % provided that t_p is less than or equal to about 4 % of t_c. Some typical value of t_c with specimens 1 mm thick is given below as:

	κ (cm^2/s)	$t_p = \ell^2/(\kappa\pi^2)$ (ms)
Aluminum	0.85	1.2
Stainless steel	0.05234	19

Typical laser pulse lengths are about ½ to 1 ms for the so-called "normal mode" lasers, and thus, with 1 mm thick specimens the technique would be fairly accurate for stainless steel but not very good for aluminum. Thicker specimens would help, but this would make the rear surface temperature rise number. If our laser pulse has, say, 20 J/cm^2 in it and we use the 10.6 μm absorptance quoted earlier for as received surfaces, the anticipated temperature rises at the back of the 1-mm specimens would be as shown below.

	A	E_0 (J/cm^2)	ρ (g/cm^3)	c_p (J/g °C)	$T_\infty = E_0/\rho c\ell$ (°C)
Aluminum	0.04	0.8	2.7	1.05	2.8
Stainless steel	0.4	8.0	8.0	0.628	16

We see the need to use shorter laser pulses with the same amount of energy. Another solution would be to carry out a more detailed heat-flow calculation. Both tailoring of the pulse shape and more detailed calculations are usually employed in current applications of laser flash techniques [51].

7.9.1.3 Analytical Solution of Heat Transfer Equation with Spatial Dependent Laser Pulse Heating

This section presents analysis of the conduction heating process introduced for a practical Nd:YAG Gaussian laser pulse with a position-dependent intensity. An analytical solution to the problem is obtained with appropriate boundary conditions for a typical Continuous Wave Laser (CW) irradiation was developed in Sect. 7.9.1.1 in above and now we pay our attention to pulse laser irradiation. When the laser intensity is rather low, no phase transition occurs and the only effect of laser absorption is heating of the material. In metals the laser radiation is absorbed by "free electron," and the energy transfer in metals is also due to electron heat conduction. The temperature field is described by the standard Fourier heat conduction and for laser pulse with a position-dependent intensity, for this form of pulse input, the Fourier differential equation is described by Eq. 7.124. $Q(x, y, z, t)$ in this equation is defined by the Drude–Zener theory which leads to the following expression [40].

$$Q(x, y, z, t) = Ae^{-\beta t}\mu I_{\max}(x, y, t)e^{-\mu z} \tag{7.179}$$

where A is the surface absorptivity, β is the time pulse parameter, μ is the absorption coefficient of the material, and $I_{\max}(x, y, t)|_{z=0}$ is the maximum laser radiation intensity at the material surface $(z = 0)$. According to Sparks [52] μ is independent

of temperature, while surface absorption coefficient of the material, $A = 1 - R$, where R is the surface reflectivity, is a linear function of the surface temperature

$$A = A_0 + A_1(T - T_0) \tag{7.180}$$

where A_0 is the surface absorptivity at room or ambient temperature, T_0, and for most of engineering materials, the absorption factor A_1 is almost unity for Nd:YAG laser wavelength. This temperature dependence of surface absorptivity results from the fact that A is proportional to the electron–phonon collision frequency which, in turn, is proportional to the crystal lattice temperature.

For metals and certain applications, at temperatures above the Debye temperature, it can be assumed that $k(T)$ and $c(T)$ do not change dramatically with temperature. Where again $c(T)$ is the density of the material of the workpiece, is the temperature dependent specific heat of the material, $k(T)$ is the temperature dependent thermal conductivity and details of these analysis were discussed in Sect. 7.7.

Therefore, assuming constant specific heat and thermal conductivity for a particular time interval, Eq. 7.124 can be simplified to

$$k\left(\frac{\partial^2 T}{\partial z^2}\right) + \mu I_{\max}(x, y, t)e^{-\mu z} = \rho c \frac{\partial T}{\partial t} \tag{7.181}$$

In many practical cases the transverse dimensions of laser focusing spot are large compared to the thickness of the heated layer, and heat conduction problem. Equation 7.181 can be considered one-dimensional and can be solved using standard methods, Carslaw and Jaeger [3].

It is unnecessary to solve for the complete pulse since the complete solution may be obtained by summation of the solutions for the individual parts of time exponential, then equation is linear. Rearrangement of Eq. 7.181 gives:

$$\frac{\partial^2 T(z,t)}{\partial z^2} + \frac{\mu I_{\max}}{k}\exp(-\mu z) = \frac{1}{\kappa}\frac{\partial T(z,t)}{\partial t} \tag{7.182}$$

where thermal diffusivity, $\kappa = k/(\rho c)$, with boundary conditions and $(z > 0, t > 0)$

$$\left.\frac{\partial T(z,t)}{\partial z}\right|_{z=0} = 0 \qquad \text{for} \quad \lim_{z\to\infty} T(z,t) = 0 \quad t > 0$$
$$T(z,t)|_{z\to\infty} = 0$$
$$T(z,0) = 0 \quad (z > 0)$$

Laplace transform can be used to solve certain type of partial differential equations with two or more independent variables. To solve Eq. 7.182 let us show how to solve in detail differential equation using Laplace transformation of one dimensional heat equation and homogeneous aspect of Eq. 7.182, given the Boundary

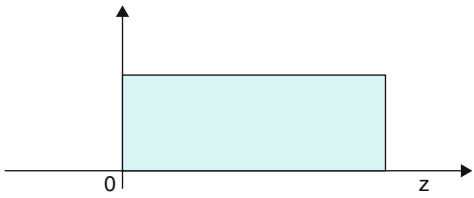


Fig. 7.39 Semi-infinite metal slab

Condition (B.C.) and Initial Condition (I.C.) as before in Sect. 7.9 and they are established below again:

$$\frac{\partial^2 T}{\partial z^2} = \frac{1}{\kappa}\frac{\partial T}{\partial t} \tag{7.183}$$

$$\text{B.C. } T(0,t) = F_0 \quad t > 0$$

$$\text{I.C. } T(z,0) = 0 \quad z > 0$$

This was done for a CW laser interaction with target materials in Sect. 7.9.1.1.

The physical model of Eq. 7.183 is a semi-infinite slab of metal with a plane face on which the origin of the z-axis is located with the positive half of the axis directed into slab. This situation is illustrated in Fig. 7.39.

The approach will be to take Laplace transform of the dependent variable $T(z,t)$ in the heat equation with respect to the time t, as a result of which an ordinary differential equation with z as its independent variable will be obtained for the transformed variable that will then depend on both the Laplace transform variables s and z. After the ordinary differential equation has been solved for the transformed variable, the inverse Laplace transform $\left(\pounds^{-1}\right)$ will be used to recover the time variation, and so to arrive at the required solution as a function of z and t. Please note that if the Laplace transform is applied to the independent variable t in the function of two variables $T(z,t)$, the variable z will behave like a constant. Consequently, the rules for transforming derivatives of functions of a single independent variable also apply to a function of two independent variables.

Using the notation $\overline{T}(z,s) = \pounds\{T(z,t)\}$ to denote the Laplace transform of $T(z,t)$ with respect to the time.

The formula for the transform of first derivative is:

$$\pounds\left\{\frac{\partial T(z,t)}{\partial t}\right\} = s\overline{T}(z,s) - T(z,0) \tag{7.184}$$

To proceed further we must now use the condition that at time $t = 0$ the material of the slab is at zero temperature, so $T(z,0) = 0$, as a result of which

$$\pounds\left\{\frac{\partial T(z,t)}{\partial t}\right\} = s\overline{T}(z,s) \tag{7.185}$$

Next, since z is regarded as a constant, one has

$$\pounds\left\{\frac{\partial^2 T(z,t)}{\partial z^2}\right\} = \frac{\partial^2 \overline{T}(z,s)}{\partial z^2} \tag{7.186}$$

Using Eq. 7.186 when taking the Laplace transform of the heat conduction equation with respect to t, and using the linearity property of the transform, one obtains

$$s\overline{T}(z,s) = \kappa\left[\frac{d^2\overline{T}(z,s)}{dz^2}\right] \quad \text{for} \quad \begin{matrix} z > 0 \\ t > 0 \end{matrix} \tag{7.187}$$

where now one can use an ordinary linear derivative with respect to z, so z can be considered to be the only independent variable. Therefore Eq. 7.187 can be rewritten as

$$\frac{d^2\overline{T}(z,s)}{dz^2} - \frac{s}{\kappa}\overline{T}(z,s) = 0 \tag{7.188}$$

Equation 7.188 has the general solution of the following with two distinctive roots for its auxiliary or characteristic equation of:

$$r^2 - \frac{s}{\kappa} = 0$$
$$r = \pm\sqrt{\frac{s}{\kappa}}$$

and

$$\overline{T}(z,s) = c_1 e^{\left(\sqrt{s/\kappa}\right)z} + c_2 e^{-\left(\sqrt{s/\kappa}\right)z} \tag{7.189}$$

Since temperature has to be finite for $z > 0$ and $t > 0$, so for $s \to +\infty$, we have $c_1 = 0$ and then the Laplace transform of temperature is seen to be given by:

$$\overline{T}(z,s) = c_2 e^{-\sqrt{s/\kappa}} = c_2 \exp\left[-\left(\sqrt{s/\kappa}\right)z\right] \tag{7.190}$$

To determine c_2 we can use boundary condition on the plane face of the slab that required $T(0,t) = F_0$ from which it follows that $\pounds\{T(0,t)\} = F_0/s$. Thus, the Laplace transform of the solution with respect to the time t is

$$\overline{T}(z,s) = \frac{F_0}{s}\exp\left[-\sqrt{\frac{s}{\kappa}}z\right] \tag{7.191}$$

So the coefficient $c_2 = F_0/s$. To recover the time variation from Eq. 7.200, it is necessary to find inverse of Laplace function $\left(\pounds^{-1}\right)$, i.e.

$$\pounds^{-1}\left\{\overline{T}(z,s)\right\} = T(z,t) \tag{7.192}$$

Result of such analysis was shown in Sect. 7.9 in more details and we just use the results here that such inverse transformation yields:

$$T(z,t) = F_0 \text{erfc}\left(\frac{z}{2\sqrt{\kappa t}}\right) \tag{7.193}$$

Where the error function, erf, is defined again as:

$$\text{erf}(z) = \frac{2}{\pi}\int_0^s e^{-u^2}du \tag{7.194}$$

and the complementary error function erfc is defined by Carslaw and Jaeger [3], as

$$\text{erfc}(z) = 1 - \text{erf}(z) \tag{7.195}$$

where z is the independent variable, and u is the dummy variable.

Continuing the forgoing mathematical calculations, the Laplace transformation of Eq. 7.182 and its related B.C. and I.C. with respect to t and substitution of boundary and initial conditions using z-direction for a one-dimensional problem yields:

$$\frac{\partial^2\overline{T}(z,s)}{\partial t^2} - g^2\overline{T}(z,s) = -\frac{I_{\max}(z,t)\exp[-\mu z]}{ks} \tag{7.196}$$

where $g^2 = s/\kappa$ and $\overline{T}(z,s)$ the transform variable, with which has a complementary and particular solution and we try to solve them here. Eq. 7.196 is in a second-order linear but nonhomogeneous differential equation, so we assume a general solution of the following type:

$$\overline{T}(z,s) = y_c + \overline{T}_p(z,s) \tag{7.197}$$

where y_c is a general solution of the homogenous part of solution to Eq. 7.196, while $\overline{T}_p(z,s)$ is the any specific function that satisfies the nonhomogeneous part of Eq. 7.196.

Solution of y_c is similar to what we found in case of Eq. 7.188 and as result Eq. 7.189, therefore the general solution to y_c with the help of its characteristic or auxiliary equation, is presented as:

$$\begin{aligned} &r^2 - g^2 = 0 \\ &r = \pm g \\ &y_c = c_1 \exp(-gz) + c_2 \exp(+gz) \end{aligned} \tag{7.198}$$

and to find the particular solution of nonhomogeneous part of Eq. 7.196, we assume the following:

$$\overline{T}_p(z,s) = Ae^{-\mu z} \tag{7.199a}$$

then

$$\frac{\partial \overline{T}_p(z,s)}{\partial z} = -A\mu e^{-\mu z} \tag{7.199b}$$

and

$$\frac{\partial^2 \overline{T}_p(z,s)}{\partial z^2} = -A\mu^2 e^{-\mu z} \tag{7.199c}$$

By substituting Eq. 7.199c into 7.196, we find that:

$$\begin{aligned} &-A\mu^2 e^{-\mu z} - g^2 A e^{-\mu z} = -\frac{I_{\max}\mu}{ks} e^{-\mu z} \\ &A = \frac{I_{\max}\mu}{ks} \frac{1}{(g^2 - \mu^2)} \end{aligned} \tag{7.200}$$

Thus, the particular solution of $\overline{T}_p(z,s)$ is given as:

$$\overline{T}(z,s) = \frac{I_{\max\mu}}{ks} \exp(-\mu z) \tag{7.201}$$

Now the summation of last term of Eqs. 7.198 and 7.201 is general solution of Eq. 7.196 and is written as:

$$\begin{aligned} \overline{T}(z,s) &= y_c + \overline{T}_p(z,s) \\ &= c_1 \exp(-gz) + c_2 \exp(+gz) + \left(\frac{I_{\max\mu}}{ks}\right)\left(\frac{1}{(\mu^2 - g^2)}\right) \exp(-\mu z) \end{aligned} \tag{7.202}$$

where c_1 and c_2 are arbitrary constants. Substituting the boundary conditions related to Eq. 7.182 in Laplace transform plane indicates that:

$$c_1 = \frac{I_{\max}\mu^2}{ksg(\mu^2 - g^2)} \quad \text{and} \quad c_2 = 0$$

Therefore, the complex solution of Eq. 7.196 in the transform plane is:

$$\overline{T}(z,s) = -\frac{I_{\max}\mu}{ks}\left[\frac{\mu\exp[-gz]}{g(\mu^2 - g^2)} - \frac{\exp[-\mu z]}{(\mu^2 - g^2)}\right] \tag{7.203}$$

The problem now is to invert the solution of Eq. 7.203 which is a product of two s-functions. There are two ways to do this. The first is a convolution integral method and the second one is to entail expansion of the functions into partial fractions. Using the second method, the full solution obtained by inverse Laplace transformation of Eq. 7.203 is

$$\begin{aligned} T(z,t) = \frac{I_{\max}\mu}{ks}\Bigg\{ & -\frac{4}{\mu}\sqrt{\frac{\kappa t}{\pi}}\exp\left(-\frac{z^2}{4\kappa t}\right) \\ & -\left(\frac{1-\mu z}{\mu^2} - \frac{1+\mu z}{\mu^2}\right)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}}\right) \\ & -\frac{1}{\mu^2}\left[\exp\left(\kappa\mu^2 - t - \mu z\right)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} - \mu\sqrt{\kappa t}\right)\right] \\ & -\frac{1}{\mu^2}\left[\exp\left(\kappa\mu^2 - t - \mu z\right)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} + \mu\sqrt{\kappa t}\right)\right] \\ & -\frac{2}{\mu^2}\exp(-\mu z)\left[1 - \exp\left(\kappa\mu^2 t\right)\right]\Bigg\} \end{aligned} \tag{7.204}$$

where

$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z) = \frac{2}{\pi}\int_z^\infty e^{-u^2}du$, defined by Eqs. 7.194 and 7.195 and knowing that:

$$\operatorname{ierfc}(z) = \int_z^\infty \operatorname{erfc}\xi d\xi = \frac{1}{\sqrt{\tau}}\exp\left(-z^2\right)\operatorname{erfc}(z) \tag{7.205}$$

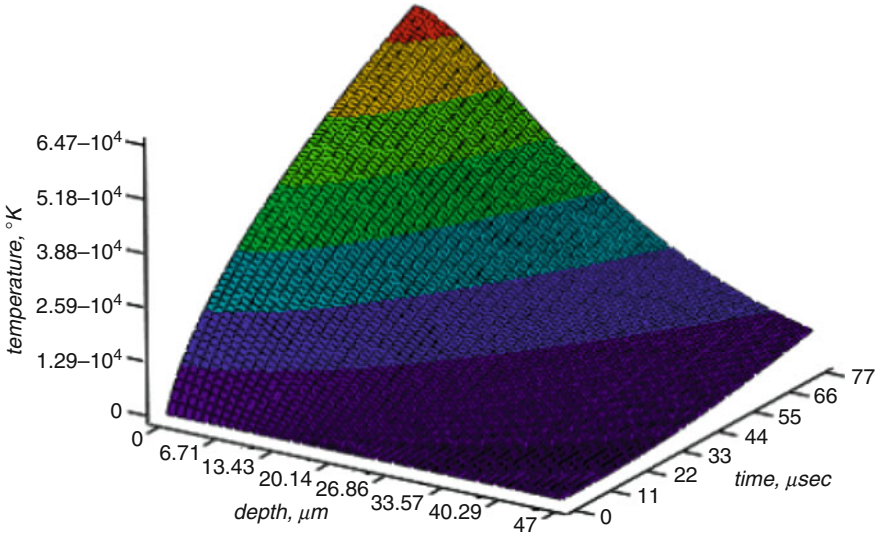


Fig. 7.40 Temperature distribution inside a material

The final solution of Eq. 7.204 is:

$$\begin{aligned} T(z,t) = {} & \frac{2I_{\max}}{k}\sqrt{\kappa t}\,\mathrm{ierfc}\left(\frac{z}{2\sqrt{\kappa t}}\right) \\ & - \frac{I_{\max}}{k\mu}\exp(-\mu z) \\ & + \frac{I_{\max}}{2k\mu}\left[\exp\left(\kappa\mu^2 - t - \mu z\right)\mathrm{erfc}\left(\mu\sqrt{\kappa t} - \frac{z}{2\sqrt{\kappa t}}\right)\right] \\ & + \frac{I_{\max}}{2k\mu}\left[\exp\left(\kappa\mu^2 - t + \mu z\right)\mathrm{erfc}\left(\mu\sqrt{\kappa t} + \frac{z}{2\sqrt{\kappa t}}\right)\right] \end{aligned} \tag{7.206}$$

Equation 7.206 gives the temperature profile inside the material for a given beam power intensity is shown in Fig. 7.40

It should be noted that as the time tends to infinity in Eq. 7.206, i.e.,

$$\lim_{t\to\infty} T(z,t) = \infty \tag{7.207}$$

no steady state solution exists for the temperature distribution.

In the majority of calculations regarding the laser heating of solids surfaces, the temperature dependence of the surface reflectivity usually is neglected. If we set $A = 0$, growth of the temperature at the surface in the center of laser beam with time may be obtained by setting $z = 0$ into Eq. 7.206, i.e.,

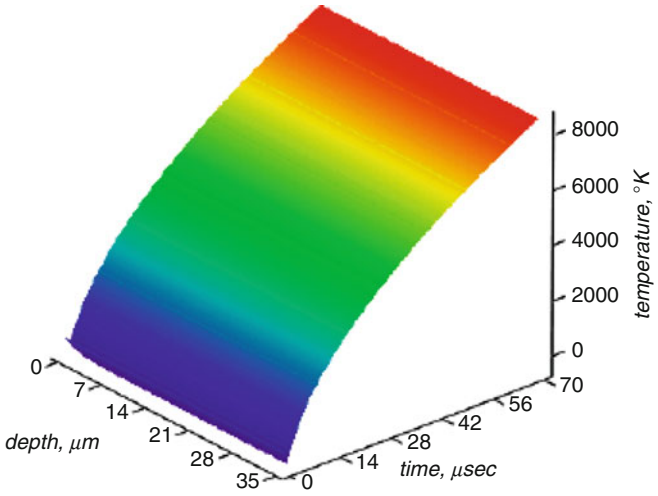


Fig. 7.41 Temperature distribution on the surface of the material

$$T(0,t) = \frac{I_{\max}}{k\mu}\left[2\frac{\sqrt{\kappa\mu^2 t}}{\pi} + \exp\left(\kappa\mu^2 t\right)\mathrm{erfc}\left(\mu\sqrt{\kappa t}\right) - 1\right] \tag{7.208}$$

Graphical representation of Eq. 7.208 is shown in Fig. 7.41.

Distribution in Fig. 7.41 represents the surface temperature of the workpiece during the laser interaction, and roughly one can estimate if there is going to be a drilling or melting on that surface. Writing in MathCAD parametric program as a function of thermal properties for different materials, (thermal conductivity of lead $k = 19.66,\ c_p = 0.151$; titanium, $k = 20.5,\ c_p = 0.782$, and 304 stainless steel $k = 41.84.\ c_p = 0.418$. Typical result of such a parametric study is illustrated in Fig. 7.42.

Using very complicated function, programming in MathCAD is not always easy. Thus, to obtained similar result, one can simply substitute values of thermo-physical parameters for solid first, then for liquid and by displaying those on two different graphs compare results. This method was used on Eq. 7.214, because equation is very complex.

Further examination of Eq. 7.206, assuming maximum absorption coefficient; let us say that μ goes to infinity, one obtains the following result.

$$T(z,t) = \frac{2I_{\max}}{k}\sqrt{\kappa t}\,\mathrm{ierfc}\left(\frac{z}{2\sqrt{\kappa t}}\right) \tag{7.209}$$

Graphical representation of Eq. 7.209 is shown in Fig. 7.43.

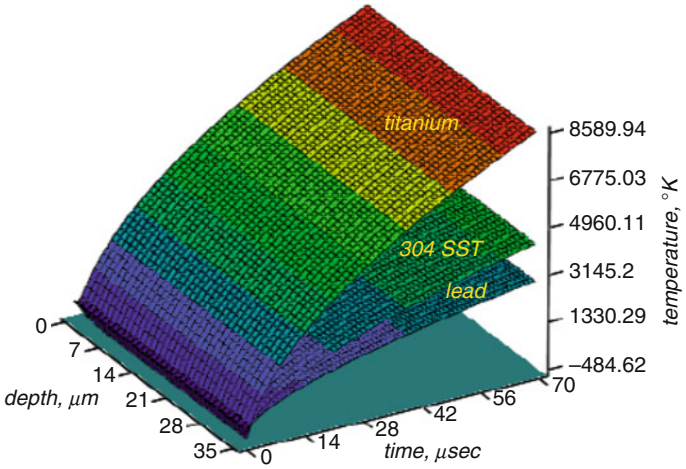


Fig. 7.42 Parametric study of thermal properties

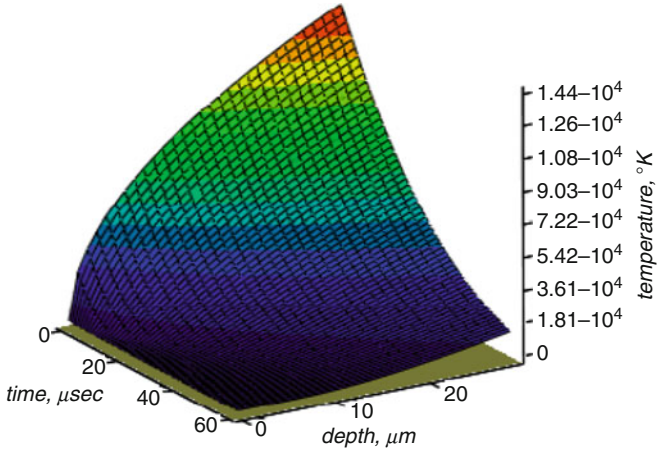


Fig. 7.43 Temperature distribution of the material with respect to time

Differentiation of Eq. 7.206 with respect to gives the temperature gradient inside the material, i.e.

$$\begin{aligned}\frac{d}{dz}T(z,t) = {} & \frac{I_{\max}}{k}\exp(-\mu z) - \frac{I_{\max}}{k}\operatorname{erf}\left(\frac{z}{2\sqrt{\kappa t}}\right) \\ & + \frac{I_{\max}}{2k}\exp\left(\kappa\mu^2 t - \mu z\right)\operatorname{erfc}\left(\mu\sqrt{\kappa t} - \frac{z}{2\sqrt{\kappa t}}\right) \\ & + \frac{I_{\max}}{2k}\exp\left(\kappa\mu^2 t + \mu z\right)\operatorname{erfc}\left(\mu\sqrt{\kappa t} - \frac{z}{2\sqrt{\kappa t}}\right) \end{aligned} \tag{7.210}$$

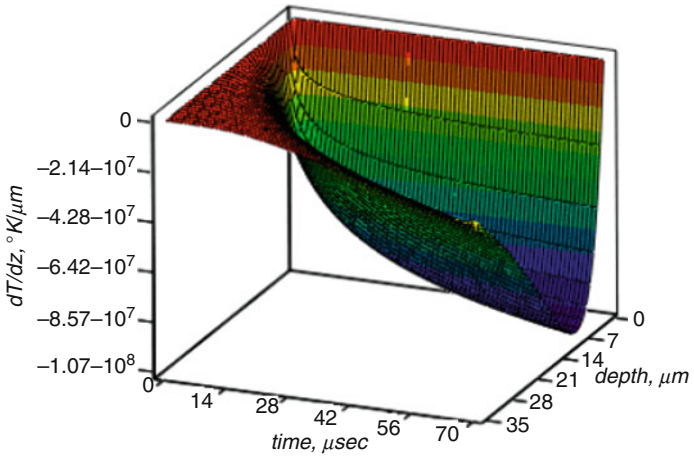


Fig. 7.44 Temperature gradient distribution inside the material with respect to time

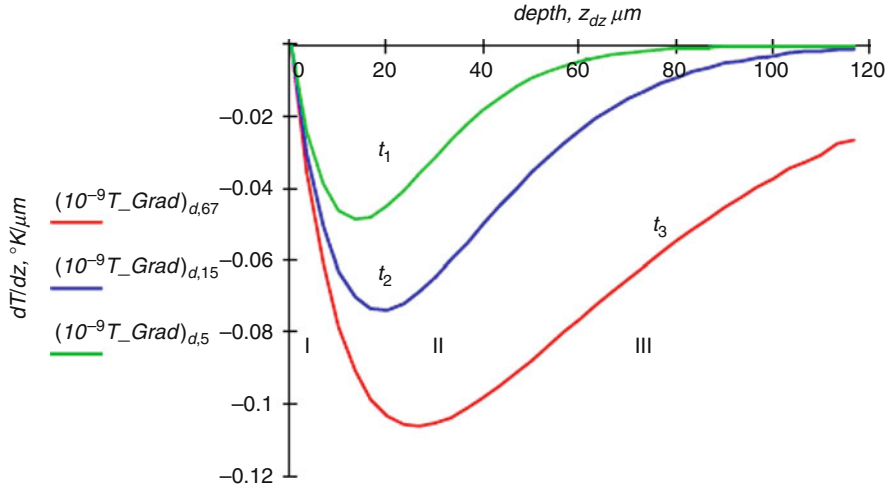


Fig. 7.45 Temperature gradient distribution inside the material for different time instances

It is evident that $dT(z,t)/dz$, Fig. 7.44 will only be zero at the surface, i.e., maximum temperature will occur at the surface.

The variation of $dT(z,t)/dz$ with distance z i.e., (depth, expressed in μm) as a function of three different time instances is shown in Fig. 7.45.

As illustrated in Fig. 7.45 the slope of the curves decreases reaching the minimum and then increases to attain almost zero as the temperature profile becomes almost asymptotic with z. In this case, the behavior of $dT(z,t)/dz$ with z may be divided into three regions, which are indicated in Fig. 7.45.

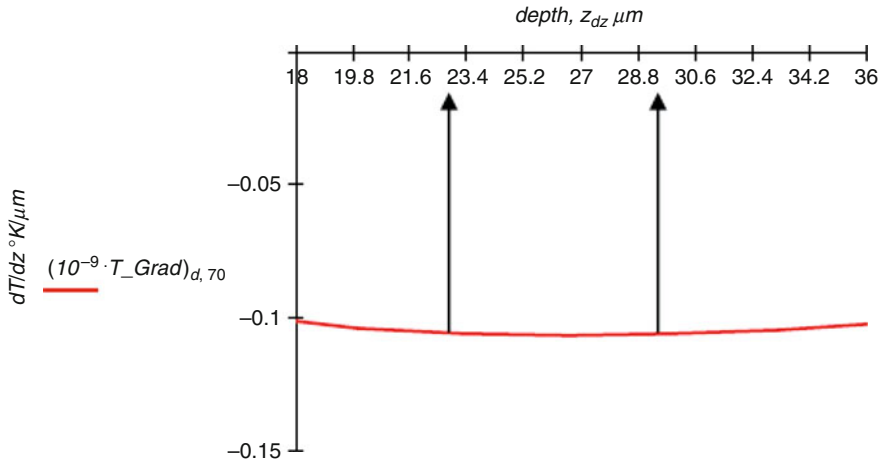


Fig. 7.46 Equilibrium distance

In the first region, the heat gain due to laser irradiation dominates the conduction losses, i.e., the internal energy increase is considerably high as compared to conduction losses. In the second region, the slope has a $z =$ minimum value; in this case, the energy gain due to incident laser beam balances the conduction losses, i.e., the internal energy of the material remains almost constant. In this case, the distance z corresponding to this point may be defined as the equilibrium distance $(z)_{\text{eq}}$, (Fig. 7.46), and $dT(z,t)/dz$ becomes $(dT(z,t)/dz)_{\min}$.

In the third region, the slope increases to reach almost zero. In this region, conduction losses are dominant and the energy gain to the external field is insignificant, i.e., the internal energy decreases as the distance increases.

Variation of the equilibrium distance show in Fig. 7.46, can be dimensionalized by:

$$(z\mu)_{\text{eq}} = C\left(\kappa\mu^2 t\right)_q^m \tag{7.211}$$

and is expected, that increase in heating time increases the dimensionless equilibrium distance, which in turn increases the dimensionless equilibrium temperature, which can be defined as:

$$\left(\frac{T(z,t)}{I_{\max}/k\mu}\right)_{\text{eq}} = C(z\mu)_q^m \tag{7.212}$$

where C is the constant and m is the power. In other words the equilibrium temperature is defined as the temperature where $dT(z,t)/dz$ is minimum.

Based on research done by Yilbas and Sami [53], the relationship between equilibrium temperature and equilibrium distance on the logarithmic scale is linear for all materials and pulse lengths.

If the pulse length is approximately 10^{-9} s the analysis of the heat transfer process using the Fourier equation becomes invalid. In this case the heating process would be nonequilibrium.

In semi-infinite medium as shown in Fig. 7.39, at $T = 0$ with no heat flow at time $t = 0$ and with the heat source defined in Eq. 7.213 below,

$$I(z) = \varepsilon I_0(0) e^{-\mu z} \tag{7.213}$$

The solution to this same problem is as:

$$\begin{aligned} T(z,t) = {} & \frac{2I_{\max}\sqrt{\kappa t}}{\mu k}\,\text{ierfc}\left(\frac{z}{2\sqrt{\kappa t}}\right) \\ & - \frac{I_{\max}}{k\mu^2}\exp(-\mu z) \\ & + \frac{I_{\max}}{2k\mu^2}\exp\left[\kappa\mu^2 t - \mu z\right]\text{erfc}\left(\mu\sqrt{\kappa t} - \frac{z}{2\sqrt{\kappa t}}\right) \\ & + \frac{I_{\max}}{2k\mu^2}\exp\left[\kappa\mu^2 t + \mu z\right]\text{erfc}\left(\mu\sqrt{\kappa t} + \frac{z}{2\sqrt{\kappa t}}\right) \end{aligned} \tag{7.214}$$

and is illustrated in Fig. 7.47. Solution in Eq. 7.214, is identical to Sparks's [54] solution for infinite thickness slab $\ell >> \delta$.

Let us consider variation of the temperature distribution with respect to the difference from the material properties variations with temperature. Analytically one can determine that based on extreme material properties for maximum and minimum values of thermal properties, known from the literature. Thermal conductivity varies between 14.9 and 16.2, for ambient temperature, T_a, and 32 for vaporization temperature, T_v, heat capacity cp varies from 500 to 824, and density ρ varies from 7870 to 8000 g/m^3. Discrepancies related to the thermo physical properties of the material could be determined by substituting the lowest and the

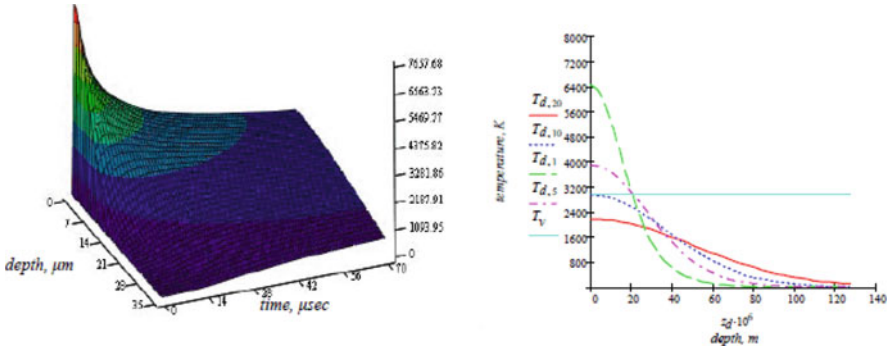


Fig. 7.47 Temperature distribution of laser irradiated material with the lowest values of thermophysical parameters, (**a**) temperature as a function of depth and time, (**b**) temperature or melting of the material with depth calculated using Eq. 7.214

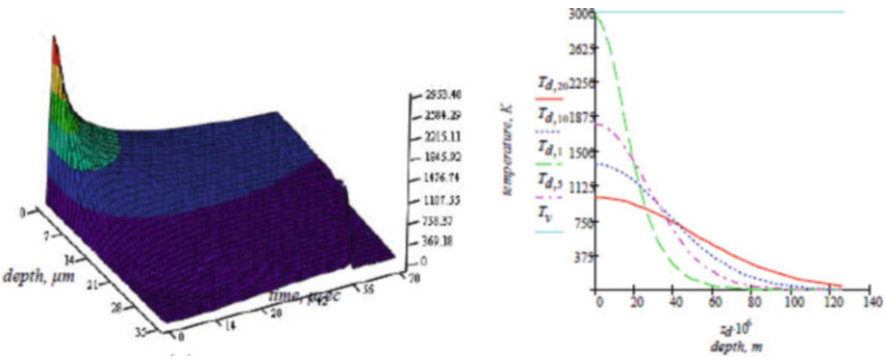


Fig. 7.48 Temperature distribution of laser-irradiated material with the pulse can be thermophysical parameters, (**a**) temperature as a function of depth and time, (**b**) temperature or melting of the material with depth

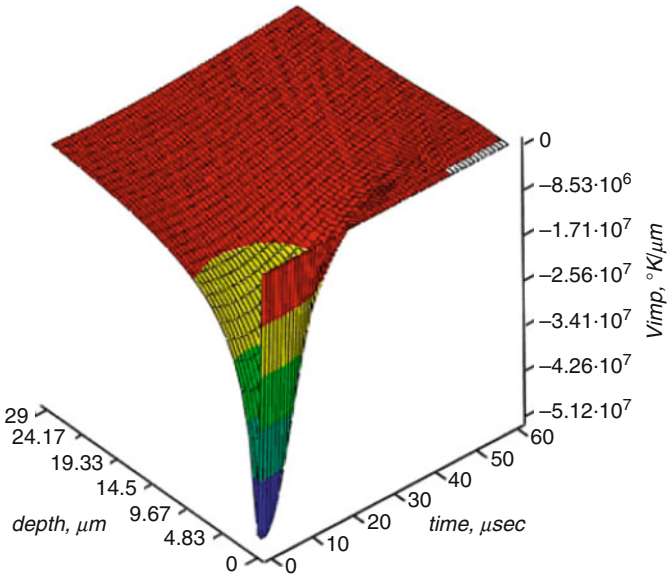


Fig. 7.49 Rate of cooling of the workpiece after the end of laser pulse

highest values to solution given by Eq. 7.214. Examining graphs in Figs. 7.47 and 7.48, one can conclude that melting temperature dependencies of the material may invoke an estimated error of about 50 %, and we talk more about melting process in Sect. 7.9.2 of this chapter.

Based on Anisimov and Khokhlov [55] rate of cooling after the end of the pulse can be estimated by Eq. 7.215 and graphically represented as in Fig. 7.49 as below.

$$V_{\text{imp}}C(z,t) = \frac{2AI_{\max}}{k}\left[\sqrt{\frac{\kappa}{t\ell_t}}\,\text{ierfc}\left(\frac{z}{2\sqrt{\kappa t}}\right) + \frac{z\text{erfc}\left(\frac{z}{2\sqrt{\kappa t}}\right)}{2t\ell_t} + \sqrt{\frac{\kappa}{t\ell_t - t_p}}\,\text{ierfc}\left(\mu\sqrt{\kappa t} - \frac{z}{2\sqrt{\kappa t}}\right) - \frac{z}{2\left(t\ell_t - t_p\right)}\text{erfc}\left(\mu\sqrt{\kappa t} + \frac{z}{2\sqrt{\kappa t}}\right)\right] \tag{7.215}$$

7.9.1.4 Analytical Solution of Heat Transfer with Time Dependent Gaussian Laser Pulse Heating

This section presents time-unsteady analysis of the conduction limited heating process introduced for a practical Nd:YAG laser pulse with a time-dependent intensity. An analytical solution to the problem is obtained with appropriate boundary conditions.

The output from a pulsed Nd:YAG laser is described by approximating the form of the true output by the subtraction of two exponential functions. This analytical form is given by:

$$I_0 = I_{\max}\left[e^{(-\beta t)} - e^{(-\gamma t)}\right] \tag{7.216}$$

For metals and certain applications, at temperatures above the Debye temperature, it can be assumed that $k(T)$ and $c(T)$ do not change with temperature. Therefore, assuming constant specific heat and thermal conductivity for a particular time interval, Eq. 7.125 can be simplified to:

$$k\left(\frac{\partial^2 T}{\partial z^2}\right) + \exp[-(\beta t + \gamma t)] \cdot \mu I_{\max}(x,y,t)_{z=0}\exp(-\mu z) = \rho c\frac{\partial T}{\partial t} \tag{7.217}$$

It is unnecessary to solve for the complete pulse, since the complete solution may be obtained by summation of the solutions for the individual parts of time exponential, then equation is linear. Rearrangement of Eq. 7.217 gives:

$$\frac{\partial^2 T}{\partial z^2} + \frac{\mu I_{\max}}{k}\exp(-\mu z)\exp[-(\beta t + \gamma t)] = \frac{1}{\kappa}\frac{\partial T}{\partial t} \tag{7.218}$$

with boundary conditions as:

$$\left.\frac{\partial T}{\partial t}\right|_{z=0} = 0 \tag{7.219}$$

$$T(\infty, t) = 0 \tag{7.220}$$

$$T(z, 0) = 0 \tag{7.221}$$

Laplace transformation of Eq. 7.218 with respect to t and substitution of boundary condition from Eq. 7.220 yields:

$$\frac{\partial^2 \overline{T}(z,s)}{\partial z^2} - g^2 \overline{T}(z,s) = \frac{-I_{\max}\mu e^{-\mu z}}{k(s+\beta)} \tag{7.222}$$

with $\overline{T}(z,s), g^2 = s/\kappa$, and s the transform variable, which has a complementary and particular solution as:

$$\overline{T}(z,s) = c_1 \exp(-gz) + c_2 \exp(gz) + \frac{I_{\max}\mu e^{-\mu z}}{k(s+\beta)(\mu^2 - g^2)} \tag{7.223}$$

where c_1 and c_2 are arbitrary constants. Substitution of boundary Equation of 7.219 and 7.221 gives these constant as, $c_2 = 0$ and

$$c_1 = \frac{I_{\max}\mu^2}{kg(s+\beta)(\mu^2 - g^2)} \tag{7.224}$$

Therefore, the complex solution in the transform plane is:

$$\overline{T}(z,s) = -\frac{I_{\max}\mu}{k(s+\beta)} \left[\frac{\mu \exp(-gz)}{g(g^2 - \mu^2)} - \frac{\exp(-\mu z)}{\left(g^2 - \mu^2\right.} \right] \tag{7.225}$$

The problem now is to invert this solution, which is a product of two p-functions. There are two ways to do this. The first is a convolution integral method and the second method is simpler, entailing expansion of the functions into partial fractions. Using the second method the full solution, obtained by inverse Laplace transformation of Eq. 7.226 is

$$T(z,t) = \frac{I_{\max}\mu}{2k}\frac{\kappa}{(\beta + \kappa\mu^2 0}$$
$$\left\{ i\mu\sqrt{\frac{\alpha}{\beta}}\exp(-\beta t)\left[\exp\left(iz\sqrt{\frac{\beta}{\kappa}}\right)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} + i\sqrt{\beta t}\right)\right]\right.$$
$$- \left[\exp\left(-iz\frac{\sqrt{\beta}}{\kappa}\right)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} - i\sqrt{\beta t}\right)\right]$$
$$+ \exp(\alpha\mu^2 t)\left[\exp(\mu z)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} + \mu\sqrt{\kappa t}\right)\right]$$
$$- \left[\exp(-\mu z)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} - \mu\sqrt{\alpha t}\right)\right]$$
$$\left. - \ 2\exp(-\mu z)[\exp(\kappa\mu^2 t) - \exp(-\beta t)]\right\} \tag{7.226}$$

Using the relationship $\operatorname{erfc}(-z) = 2 - \operatorname{erfc}(z)$, give the solution in following form:

$$T(z,t) = \frac{I_{\max}\mu}{2k}\frac{\kappa}{(\beta + \kappa\mu^2 0}$$
$$\left\{ i\mu\sqrt{\frac{\alpha}{\beta}}\exp(-\beta t)\left[\exp\left(iz\sqrt{\frac{\beta}{\kappa}}\right)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} + i\sqrt{\beta t}\right)\right.\right.$$
$$- \exp\left(-iz\frac{\sqrt{\beta}}{\kappa}\right)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} - i\sqrt{\beta t}\right)\Big]$$
$$+ \exp(\alpha\mu^2 t)\left[\left(\exp(\mu z)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} + \mu\sqrt{\kappa t}\right)\right)\right.$$
$$- \exp(-\mu z)\operatorname{erfc}\left(\frac{z}{2\sqrt{\kappa t}} - \mu\sqrt{\alpha t}\right)$$
$$\left.\left. -2\exp[-(\beta t + \mu z)]\right]\right\} \tag{7.227}$$

Note that:

$$\mathbf{i}^{2m}\mathrm{erfc}(-z) = -\mathbf{i}^{2m}\mathrm{erfc}(z) + \sum_{q=0}^{m} \frac{z^{2q}}{2^{2(m-q)-1}(2q)!(m-q)!}$$

$$\mathbf{i}^{2m+1}\mathrm{erfc}(-z) = \mathbf{i}^{2m+1}\mathrm{erfc}(z) + \sum_{q=0}^{m} \frac{z^{2q+1}}{2^{2(m-q)-1}(2q+1)!(m-q)!}$$

The temperature distribution calculated in Eq. 7.227, is illustrated in Fig. 7.50 (Fig. 7.51).

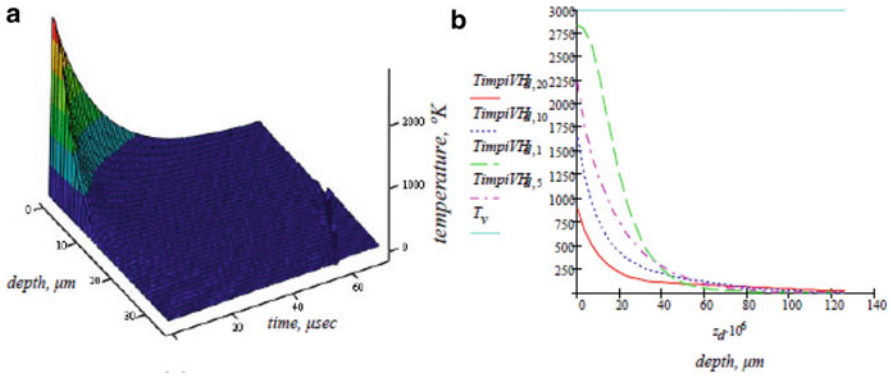


Fig. 7.50 (**a**) Temperature distribution inside the material (z-axis) with respect to time: (**b**) temperature vs. depth of the workpiece, at few different time instances from the beginning of the pulse, with the highest values of thermophysical parameters calculated using Eq. 7.227

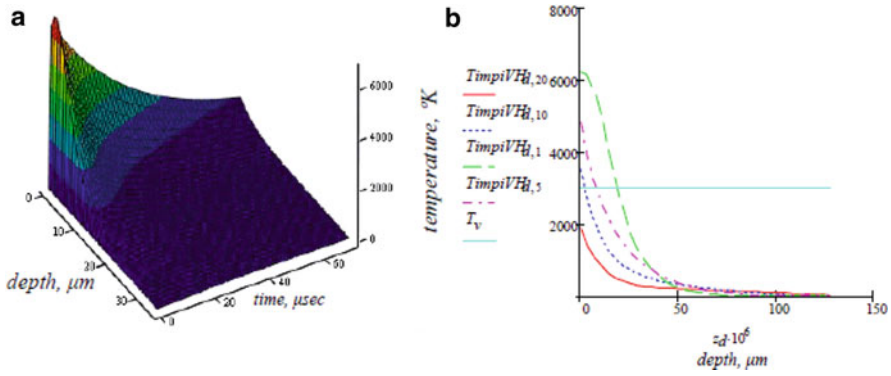


Fig. 7.51 (**a**) Temperature distribution inside the material (z-axis) with respect to time: (**b**) temperature vs. depth of the workpiece, at few different time instances from the beginning of the pulse, with the highest values of thermophysical parameters calculated using Eq. 7.227

The time at which the maximum value of the pulse occurs can be obtained by differentiating Eq. 7.227 with respect to t for the pulse. Mathematically speaking one has to find t, in the condition that the first derivative of this function should be equal to zero, which gives the condition for the maximum time $t_{\max}$.

$$t_{\max} = \frac{\ln\left(\frac{\gamma}{\beta}\right)}{\gamma - \beta} \tag{7.228}$$

or

$$\sqrt{\beta t_{\max}} = \sqrt{\frac{\ln\left(\frac{\gamma}{\beta}\right)}{\frac{\gamma}{\beta} - 1}} \tag{7.229}$$

Insertion of Eq. 7.229 into 7.227 gives the maximum temperature when the maximum pulse occurs. Solving Eqs. 7.227 and 7.229 give the condition that the maximum temperature may coincides with the maximum pulse amplitude, i.e., at this point

$$\sqrt{\beta t_{\max}} = 1.282 \quad \text{and} \quad \sqrt{\gamma/\beta} = 1.596$$

The time at which the maximum temperature occurs is a function of pulse parameters, provided that the pulse rise times are much greater than the equilibrium time $C/\kappa\mu^2$, described in Sect. 7.9.1.1. The surface temperature depends on absorption depth in the initial stages of the pulse and ability to absorb the laser depends on the shape of the pulse Yilbas et al. [41].

The knowledge of pulse parameters β and γ allows at least theoretically, to adjust the pulse for maximum desired effect based on deductions from the physics of lasing system

7.9.1.5 Analytical Solution of Heat Transfer with Time Dependent Gaussian Laser Pulse Heating with Convective Boundary Conditions

This section presents time and unsteady analysis of the conduction with limited heating process introduced for a practical Nd-YAG laser pulse with a time-dependent intensity. An analytical solution to the problem is obtained with appropriate boundary conditions.

The output from a pulsed Nd-YAG laser is described by approximating the form of the true output by the subtraction of two exponential functions. This analytical form is given by:

$$I = AI_{\max}[\exp(-\beta t) - \exp(-\gamma t)] \tag{7.230}$$

For metals and certain applications, at temperatures above the Debye temperature, it can be assumed that $k(T)$ and $c(T)$ do not change with temperature. Therefore, assuming constant specific heat and thermal conductivity for a particular time interval, Eq. 7.125 can be simplified to:

$$\rho c \frac{\partial T(z,t)}{\partial t} = k\left(\frac{\partial^2 T(z,t)}{\partial z^2}\right) + \exp(-\beta t) - \exp(-\gamma t)\left[A\mu I_{\max}(x,y,t)_{z=0}\exp(-\mu z)\right] \tag{7.231}$$

It is unnecessary to solve for the complete pulse, since the complete solution may be obtained by summation of the solutions for the individual parts of time exponential, then equation is linear. Rearrangement of Eq. 7.231 gives:

$$\frac{\partial^2 T}{\partial z^2} + A\frac{\mu I_{\max}}{k}\exp(-\beta t) - \exp(-\gamma t)\exp(\mu z) = \frac{1}{\kappa}\frac{\partial T}{\partial t} \tag{7.232}$$

with boundary conditions

$$\left.\frac{\partial T}{\partial t}\right|_{z=0} = \frac{h}{k}[T(0,t) - T_0] \tag{7.233}$$

$$T(\infty, t) = 0 \tag{7.234}$$

$$T(z,0) = 0 \tag{7.235}$$

Complete solution can be obtained by subtraction of solutions for the individual parts of the time exponential pulse. It should be noted that for the solution of a complete pulse, the ambient temperature is considered as zero ($T_0 = 0$). Hence, the heat transfer for the half-pulse becomes

$$\frac{\partial^2 T}{\partial z^2} + A\frac{\mu I_{\max}}{k}\exp\left[-(\beta t + \mu z\right] = \frac{1}{\kappa}\frac{\partial T}{\partial t} \tag{7.236}$$

The solution of Eq. 7.236 can be obtained through the Laplace transformation method, with respect to t as:

$$\frac{\partial^2 \overline{T}(z,s)}{\partial z^2} + \frac{AI_{\max}\mu\exp(-\mu z)}{k(s+\beta)} = \frac{1}{\kappa}\left[s\overline{T}(z,s) - T(z,0)\right] \tag{7.237}$$

with $\overline{T}(z,s)$, and $T(z,0) = 0$, where s the transform variable, which has a complementary and particular solution as before.

The full solution, obtained by inverse Laplace transformation of Eq. 7.237, is

$$T(z,t) = a_{10}\left\{\begin{array}{l} i\mu\sqrt{\dfrac{\kappa}{\beta}}\exp(-\beta t)\left[\begin{array}{l} l_1\exp\left(iz\sqrt{\dfrac{\beta}{\kappa}}\right)\mathrm{erfc}\left(\dfrac{z}{2\sqrt{\kappa t}}+i\sqrt{\beta t}\right) \\ l_2\exp\left(-iz\sqrt{\dfrac{\beta}{\kappa}}\mathrm{erfc}\left(\dfrac{z}{2\sqrt{\kappa t}}-i\sqrt{\beta t}\right)\right)\end{array}\right] \\ +\mu\sqrt{\kappa}\exp(\alpha\mu^2 t)\left[\begin{array}{l} l_3\exp(-\mu z)\mathrm{erfc}\left(\dfrac{z}{2\sqrt{\kappa t}}-\mu\sqrt{\beta t}\right) \\ -l_4\exp(\mu z)\mathrm{erfc}\left(\dfrac{z}{2\sqrt{\kappa t}}+\mu\sqrt{\beta t}\right)\end{array}\right] \\ -w_1 l_5\exp\left(w_1\dfrac{z}{\sqrt{\kappa}}\right)\exp\left(w_1^2\right)\mathrm{erf}\left(\dfrac{z}{2\sqrt{\kappa t}}+w_1\sqrt{t}\right)\end{array}\right\} \\ +\left\{\begin{array}{l} \dfrac{a_{20}}{\beta+\kappa\mu^2}\exp\left(\alpha\mu^2 t\right)-\exp(-\beta t)\exp(-\mu z) \\ -\dfrac{a_{30}}{w_1}\left[\begin{array}{l} -\exp\left(w_1\dfrac{z}{\sqrt{\kappa}}\right)\exp\left(w_1^2 t\right)\mathrm{erf}\left(\dfrac{z}{2\sqrt{\kappa t}}+w_1\sqrt{t}\right) \\ \mathrm{erfc}\left(\dfrac{z}{2\sqrt{\kappa t}}\right)\end{array}\right]\end{array}\right\} \tag{7.238}$$

where

$$a_{10} = \frac{-\mu I_{\max}\kappa(h+k\mu)}{k^2} \tag{7.239}$$

$$a_{20} = \frac{I_{\max}\kappa\mu}{k} \tag{7.240}$$

$$a_{30} = \frac{\sqrt{\kappa}h}{k}T_0 \tag{7.241}$$

The temperature distribution calculated in Eq. 7.238 is illustrated in Fig. 7.52 (Figs. 7.53, 7.54, and 7.55).

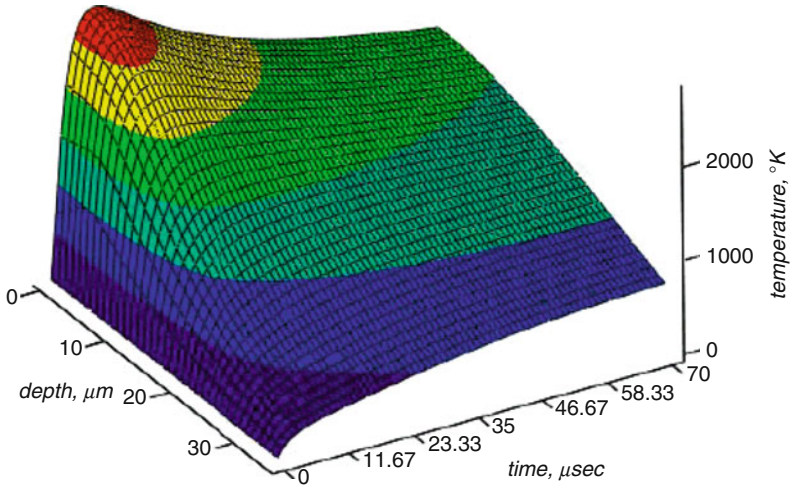


Fig. 7.52 (**a**) Temperature distribution inside the material (z-axis) with respect to time: (**b**) temperature vs. depth of the workpiece, at few different time instances from the beginning of the pulse, with the highest values of thermophysical parameters calculated using Eq. 7.238

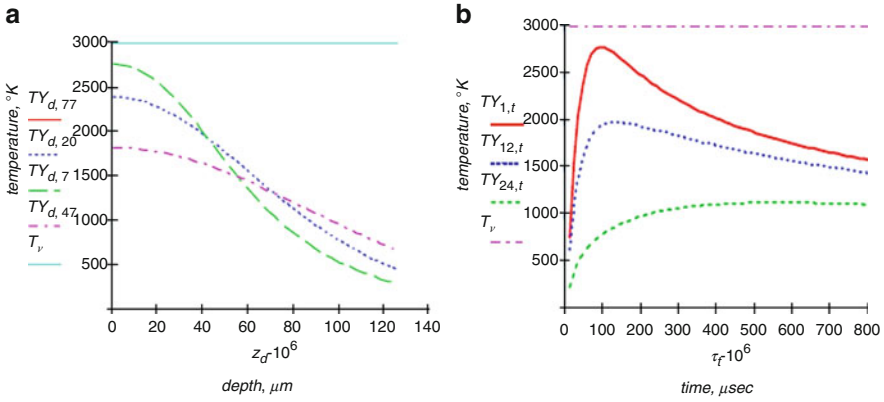


Fig. 7.53 (**a**) Temperature distribution inside the material (z-axis) for few different time instances; (**b**) transient distribution of the workpiece, at few different depths, with the highest values of thermophysical parameters calculated using Eq. 7.238

7.9.1.6 Analytical Solution of Governing Heat Transfer Equation, Considering Transfer Evaporative Case

This section presents analysis of the heating process. In the heating analysis, evaporation is considered as the exclusive phenomenon taking place during the ablation process. Based on, forgoing discussion, melting without vaporization can be produced only in very narrow range of laser parameters. Usually both phase

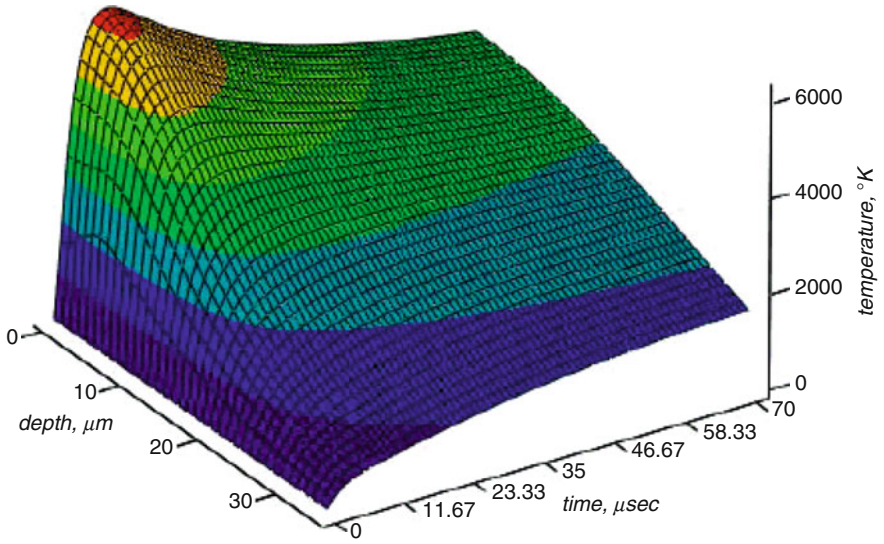


Fig. 7.54 Temperature distribution inside the material (z-axis) with respect to time, with the lowest values of thermophysical parameters calculated using Eq. 7.238

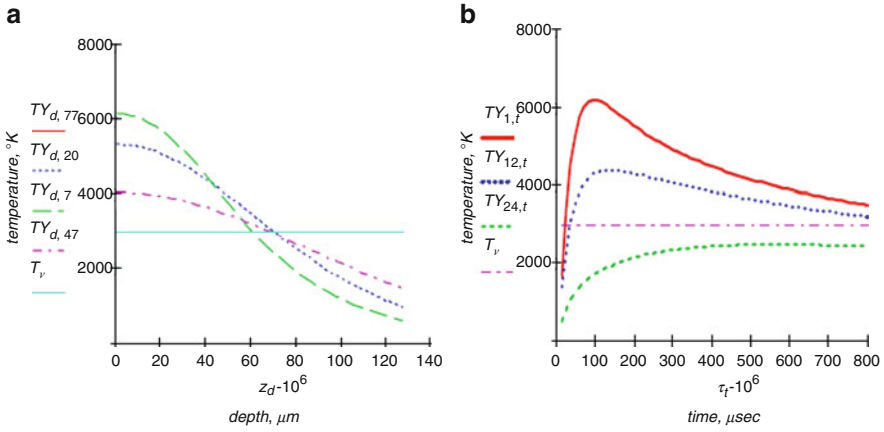


Fig. 7.55 (**a**) Temperature distribution inside the material (z-axis) for few different time instances; (**b**) transient distribution of the workpiece, at few different depths, with the lowest values of thermophysical parameters calculated using Eq. 7.238

transitions (melting and vaporization) occurs almost simultaneously in standard laser applications. Note that the latent heat of vaporization is larger than that of fusion by the factor of 20-50 times.

Thus, evaporation plays the most important part in the energy balance. To simplify our discussion we will assume that the vaporization occurs in vacuum and the vapor does not absorb the laser energy. When the vapor does not absorb the

laser light, the process in the condensed and gaseous phases can be considered separately. An analytical solution to the problem is obtained with appropriate boundary conditions. In this case, plasma formation, liquid expulsion, and nucleate boiling are omitted.

Solid or liquid evaporates at any temperature greater than 0°K. The evaporation rates strongly depend on the temperature, T. This dependence can be written in the form

$$V_{\mathrm{n}} = V_0 \exp\left(-\frac{U}{T}\right) \tag{7.242}$$

where V_{n} is the normal component of the evaporation front velocity, and

$$U = \frac{MH_{\mathrm{v}}}{N_{\mathrm{a}} k_{\mathrm{B}}} \tag{7.243}$$

and H_{v} is the latent heat of vaporization per unit mass, k_{B} is Boltzmann's constant. M is the atomic mass, and V_0 is a constant whose value is the order of the speed of sound in the condensed phase and finally N_{a} is Avogadro's number. The Fourier differential equation allowing a phase change process can be written as:

$$\rho(T)\frac{\partial}{\partial t}\int_{T_0}^{T} c(T)dT = \frac{\partial}{\partial x}\left[k(T)\frac{\partial T}{\partial x}\right] + \frac{\partial}{\partial y}\left[k(T)\frac{\partial T}{\partial y}\right] + \frac{\partial}{\partial z}\left[k(T)\frac{\partial T}{\partial z}\right] + [\rho(T)c(T)]V\frac{\partial T}{\partial z} + Q \tag{7.244}$$

where $\rho(T)$ is the density of the material of the workpiece, $c(T)$ is the temperature dependent specific heat of the material, $k(T)$ is the temperature dependent thermal conductivity, V is the recession velocity which is the velocity of the liquid–vapor interface, and $T = T(x, y, z)$ is the resulting three-dimensional time dependent distribution in the material, t is time T_0 is the initial temperature, and x, y, z are the spatial coordinates.

$Q = Q(x, y, z, t)$ is the rate at which heat is supplied to the solid per unit time per unit volume, depends on the laser pulse parameters and optical properties of target material irradiated and defined by 7.179.

Sparks [54] has defined absorption coefficient of the material as a linear function of the surface temperature in Eq. 7.151. For practical reason the absorption factor A we assume as unity. This temperature dependence of A results from the fact that A is proportional to the electron–phonon collision frequency which, in turn, is proportional to the crystal lattice temperature.

The recession velocity of the surface can be formulated from energy balance at the free surface of irradiated workpiece. In this case the energy flux at the free surface can be written as

$$I_0 = \rho(T)V[c(T)T_s - H_{\mathrm{v}}] \tag{7.245}$$

Where

$$I_0 = I_{\max}(1 - R) \tag{7.246}$$

and T_s is the surface temperature and H_v is latent heat of vaporization and R is surface reflectivity. Rearranging of Eq. 7.245 yields

$$V = \frac{I_0}{\rho(T)[c(T)T_s - H_v]} \tag{7.247}$$

For metals and certain applications, at temperatures above the Debye temperature, it can be assumed that $k(T)$, $\rho(T)$ and $c(T)$ do not change dramatically with temperature.

Therefore, assuming constant specific heat and thermal conductivity for a particular time interval, Eq. 7.244 can be simplified to

$$\rho c \frac{\partial T}{\partial t} = k\left(\frac{\partial^2 T}{\partial z^2}\right) + \rho c V \frac{\partial T}{\partial z} + I_{\max}(x, y, t)_{z=0}(1 - R)\mu e^{-(\mu z + \beta t)} \tag{7.248}$$

It should be noted that peak power intensity, $I_{\max}$ does not vary with time. Since the surface temperature is time dependant, the recession velocity varies with time. This result is nonlinear form of Eq. 7.247, which cannot be solved analytically by Laplace transformation method.

For a known surface temperature there is a unique value for recession velocity, so an iterative method can be applied here to solve Eq. 7.248 analytically. In our case keeping the recession velocity constant in Eq. 7.247 enables us to determine the surface temperature analytically and after obtaining the surface temperature the recession velocity can be calculated using Eq. 7.247. This procedure can be repeated so long as the surface temperature and recession velocity converge to correct results. Rearrangement of Eq. 7.248 gives

$$\frac{\partial^2 T}{\partial z^2} + \frac{V}{\kappa}\frac{\partial T}{\partial z} + \frac{\mu I_0}{k} e^{-\beta t} e^{-\mu z} = \frac{1}{\kappa}\frac{\partial T}{\partial t} \tag{7.249}$$

where thermal diffusivity $\kappa = \frac{k}{\rho c}$, with boundary conditions

$$\begin{cases} \frac{\partial T}{\partial t}\big|_{z=0} = \dfrac{\rho V L}{k} \\ T(z,t)|_{z\to\infty} = 0 \\ T(z,t)|_{t=0} = 0 \end{cases} \tag{7.250}$$

We can apply the Laplace transform, defined as:

$$\overline{T}(z,s) = \int_0^\infty T(z,t)e^{-st}dt \tag{7.251}$$

Laplace transformation of Eq. 7.249 with respect to t and substitution of boundary conditions from Eq. 7.250, gives:

$$\frac{\partial^2\overline{T}(z,s)}{\partial z^2} + \frac{V}{\kappa}\frac{\partial\overline{T}(z,s)}{\partial z} - \frac{s}{\kappa}\overline{T}(z,s) = -\frac{I_0\mu}{k(s+\beta)}e^{(-\mu z)} \tag{7.252}$$

with $\overline{T}(z,s)$, and s the transform variable in Laplace plane, which has a complementary (homogeneous) and particular (nonhomogeneous) solution as:

$$\overline{T}(z,s) = \overline{T}(z,s)_c + \overline{T}(z,s)_p \tag{7.253}$$

The homogeneous part of Eq. 7.253, $\overline{T}(z,s)_c$ is as:

$$\frac{\partial^2\overline{T}(z,s)}{\partial z^2} + \frac{V}{\kappa}\frac{\partial\overline{T}(z,s)}{\partial z} - \frac{s}{\kappa}\overline{T}(z,s) = 0 \tag{7.254}$$

The characteristic or auxiliary equation for the homogeneous solution can be written as:

$$r^2 + \frac{V}{\kappa}r - \frac{s}{\kappa} = 0 \tag{7.255}$$

which Eq. 7.255 has two distinctive roots as:

$$r_{1,2} = -\frac{V}{2\kappa} \pm \frac{\sqrt{V^2+4s\kappa}}{2\kappa} \tag{7.256}$$

The homogeneous solution from Eq. 7.255, then results in:

$$\overline{T}(z,s)_c = c_1e^{r_1z} + c_2e^{r_2z} \tag{7.257}$$

or

$$\overline{T}(z,s)_c = e^{-\frac{Vz}{2\kappa}}\left[c_1e^{-\left(\frac{\sqrt{V^2+4s\kappa}}{2\kappa}\right)z} + c_2e^{+\left(\frac{\sqrt{V^2+4s\kappa}}{2\kappa}\right)z}\right] \tag{7.258}$$

For the particular solution, one can provide exponential solution as:

$$\overline{T}(z,s)_p = P_0e^{-\delta z} \tag{7.259}$$

where P_0 is some arbitrary constant coefficient. Substitution of $\overline{T}(z,s)_p$ given by Eq. 7.258 into Eq. 7.252 results in:

$$P_0\delta^2 e^{-\delta z} - \frac{V}{\kappa}P_0\delta e^{-\delta z} - \frac{s}{\kappa}P_0\delta e^{-\delta z} = H_0 e^{-\delta z} \tag{7.260}$$

where

$$H_0 = I_0\frac{1}{(s+\beta)}\frac{\delta}{k} \tag{7.261}$$

and

$$P_0 = -\frac{H_0\kappa}{s + V\delta - \kappa\delta^2} \tag{7.262}$$

and substituting Eq. 7.262 into Eq. 7.259, and the end result of that along with Eq. 7.258 into Eq. 7.253, we obtain:

$$\overline{T}(z,s) = e^{-\frac{Vz}{2\kappa}}\left[c_1 e^{-\left(\frac{\sqrt{V^2+4s\kappa}}{2\kappa}\right)z} + c_2 e^{+\left(\frac{\sqrt{V^2+4s\kappa}}{2\kappa}\right)z}\right] - \frac{H_0\kappa}{s + V\delta - \kappa\delta^2}e^{-\delta z} \tag{7.263}$$

Now defining:

$$V_1 = V\delta - \kappa\delta^2 \tag{7.264}$$

and substituting Eq. 7.264 and using boundary Eq. 7.250, then $c_2 = 0$ and all into Eq. 7.263, we get:

$$\overline{T}(z,s) = c_1 e^{\frac{1}{2\kappa}\left[V+\sqrt{V^2+4s\kappa}\right]z} + \frac{I_0\delta\kappa}{k}\frac{e^{-\delta z}}{(s+\beta)(S+V_1)} \tag{7.265}$$

We can calculate c_1 by using boundary condition of Eq. 7.250 in Laplace plane and defining:

$$\varpi_1 = \frac{I_0\delta\kappa}{k} \tag{7.266}$$

$$\begin{aligned}\frac{\overline{T}(z,s)}{\partial z} &= \left\{-\frac{1}{2\kappa}\left[V + \sqrt{V^2+4s\kappa}\right]c_1 e^{\frac{1}{2\kappa}\left[V+\sqrt{V^2+4s\kappa}\right]z} - \frac{\delta\varpi_1\kappa e^{-\delta z}}{(s+\beta)(S+V_1)}\right\}_{z=0} \\ &= \frac{\rho VL}{ks}\end{aligned} \tag{7.267}$$

From Eq. 7.267, c_1 is going to be

$$c_1 = -\frac{2\delta\varpi_1\kappa}{(s+\beta)(s+V_1)\left(V+\sqrt{V^2+4s\kappa}\right)} - \frac{2\kappa\frac{\rho VL}{k}}{s\left(V+\sqrt{V^2+4s\kappa}\right)} \tag{7.268}$$

Hence,

$$\begin{aligned}\overline{T}(z,s) = &\frac{2\delta\varpi_1\kappa e^{-\frac{1}{2\kappa}\left[V+\sqrt{V^2+4s\kappa}\right]}}{(s+\beta)(s+V_1)\left(V+\sqrt{V^2+4s\kappa}\right)} \\ &+ \frac{\varpi_1 e^{-\delta z}}{(s+\beta)(s+V_1)} + \overline{T}_{mm}(z,s)\end{aligned} \tag{7.269}$$

where

$$\overline{T}_{mm}(z,s) = -2\kappa\rho VL\frac{e^{-\frac{1}{2\kappa}\left[V+\sqrt{V^2+4s\kappa}\right]}}{ks\left(V+\sqrt{V^2+4s\kappa}\right)} \tag{7.270}$$

For the purpose of easier handling calculation for the future, let us define: $H_1(s)$, $H_2(s)$, and $H_3(s)$ to be:

$$H_1(s) = \frac{e^{-\frac{1}{2\kappa}\left[V+\sqrt{V^2+4s\kappa}\right]}}{(s+\beta)(s+V_1)\left(V+\sqrt{V^2+4s\kappa}\right)} \tag{7.271}$$

$$H_2(s) = \frac{1}{(s+\beta)(s+V_1)} = \frac{1}{V_1-\beta}\left[\frac{1}{s+\beta} - \frac{1}{s+V_1}\right] \tag{7.272}$$

$$H_3(s) = \frac{e^{-\frac{1}{2\kappa}\left[V+\sqrt{V^2+4s\kappa}\right]}}{s\left(V+\sqrt{V^2+4s\kappa}\right)} \tag{7.273}$$

And term of these defined variables Eq. 7.269 becomes as:

$$\overline{T}(z,s) = -2\kappa\mu\varpi_1 H_1(s) + \varpi_1 e^{-\mu z}H_2(s) - 2\kappa\frac{\rho VL}{k}H_3(s) \tag{7.274}$$

Consequently the solution to Eq. 7.249 is going to be:

$$T(z,t) = \pounds^{-1}\{\overline{T}(z,s)\} \tag{7.275}$$

or

$$\begin{aligned} T(z,t) = &-2\kappa\mu\varpi_1\pounds^{-1}\{H_1(s)\} + \varpi_1 e^{-\mu z}\pounds^{-1}\{H_2(s)\} \\ &- 2\kappa\frac{\rho VL}{k}\pounds^{-1}\{H_3(s)\} \end{aligned} \tag{7.276}$$

To obtain the inverse transformation of functions $\pounds^{-1}\{H_1(s)\}$ and $\pounds^{-1}\{H_3(s)\}$, let us introduce

$$p = V^2 + 4\kappa s \quad \text{or} \quad dp = 4\kappa ds \quad \text{and} \quad s = \frac{1}{4\kappa}(p - V^2)$$

Therefore

$$\pounds^{-1}\{H_1(s)\} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{ts}H_1(s)ds \tag{7.277}$$

or

$$\pounds^{-1}\{H_1(s)\} = \frac{1}{4\kappa}e^{-\left(\frac{Vz}{2\kappa}+\frac{V^2}{4\kappa}\right)}\frac{1}{2\pi i}\int_{\overline{c}-i\infty}^{\overline{c}+i\infty}\frac{e^{-\left(\frac{Vz}{2\kappa}+\frac{V^2}{4\kappa}\right)}}{\left(\frac{p}{4\kappa}-\frac{V^2}{4\kappa}+\beta\right)\left(\frac{p}{4\kappa}-\frac{V^2}{4\kappa}+V_1\right)\left(V+\sqrt{p}\right)}dp \tag{7.278}$$

where

$$\overline{c} = 4\kappa c + V_2 \tag{7.279}$$

We can use one more transformation by letting:

$$\vartheta = \frac{1}{4\kappa}p \quad \text{or} \quad dp = 4\kappa d\vartheta \tag{7.280}$$

After tedious and lengthy algebra, we obtain

$$\pounds^{-1}\{H_1(s)\} = \frac{1}{4\kappa}e^{-\left(\frac{Vz}{2\kappa}+\frac{V^2}{4\kappa}\right)}\frac{1}{2\pi i}\int_{\hat{c}-i\infty}^{\hat{c}+i\infty}\frac{4\kappa e^{-\left(\frac{z\sqrt{s}}{\sqrt{\kappa}}\right)}e^{\vartheta t}}{\left(\vartheta-\frac{V^2}{4\kappa}+\beta\right)\left(\vartheta-\frac{V^2}{4\kappa}+V_1\right)(V+2\sqrt{\kappa}\sqrt{s})}d\vartheta \tag{7.281}$$

Where

$$\hat{c} = \overline{c}/4\kappa \tag{7.282}$$

Hence,

$$\pounds^{-1}\{H_1(s)\} = 4\kappa e^{-\frac{V}{2\kappa}\left(z+t\frac{V}{2}\right)}\pounds^{-1}\{H_4(s)\} \tag{7.283}$$

where

$$H_4(s) = \frac{4\kappa e^{-\left(\frac{z\sqrt{s}}{\sqrt{\kappa}}\right)}e^{\vartheta t}}{\left(\vartheta - \frac{V^2}{4\kappa} + \beta\right)\left(\vartheta - \frac{V^2}{4\kappa} + V_1\right)(V + 2\sqrt{\kappa}\sqrt{s})} \tag{7.284}$$

Similarly, $\pounds^{-1}\{H_3(s)\}$ can be obtained, i.e., as:

$$\pounds^{-1}\{H_3(s)\} = 4\kappa e^{-\frac{V}{2\kappa}\left(z+t\frac{V}{2}\right)}\pounds^{-1}\{H_7(s)\} \tag{7.285}$$

where

$$H_7(s) = \frac{e^{-\left(\frac{z\sqrt{s}}{\sqrt{\kappa}}\right)}}{\left(\sqrt{\kappa}\sqrt{s} - V^2\right)\left(2\sqrt{\kappa}s + V\right)^2} \tag{7.286}$$

or

$$H_7(s) = \frac{e^{-\left(\frac{z\sqrt{s}}{\sqrt{\kappa}}\right)}}{8\kappa\sqrt{\kappa}(\sqrt{s} - \xi)(\sqrt{s} + \xi)^2} \tag{7.287}$$

where

$$\xi = \frac{V}{2\sqrt{\kappa}} \tag{7.288}$$

Introducing partial fraction and rearrangement yields:

$$H_7(s) = \frac{1}{8V^2\sqrt{\kappa}}\left[\frac{e^{-\left(\frac{z\sqrt{s}}{\sqrt{\kappa}}\right)}}{(\sqrt{s} - \xi)} - \frac{e^{-\left(\frac{z\sqrt{s}}{\sqrt{\kappa}}\right)}}{(\sqrt{s} + \xi)} - \frac{e^{-\left(\frac{z\sqrt{s}}{\sqrt{\kappa}}\right)}}{\sqrt{\kappa}(\sqrt{s} + \xi)^2}\right] \tag{7.289}$$

It is noted from the Laplace inversion that

$$\pounds^{-1}\left\{\frac{e^{-k\sqrt{s}}}{\sqrt{s}+s}\right\} = \frac{1}{\sqrt{\pi t}}e^{-\frac{k^2}{2t}} - ae^{ak}e^{a^2t}\text{erfc}\left(a\sqrt{t}+\frac{k}{2\sqrt{t}}\right) \tag{7.290}$$

where complementary error function is defined as:

$$\text{erfc}(z) = 1 - \text{erf}(z) = \frac{2}{\pi}\int_z^{\infty} e^{-u^2}du \tag{7.291}$$

Therefore

$$\pounds^{-1}\left\{\frac{e^{-\frac{z}{\sqrt{\kappa}}\sqrt{s}}}{\sqrt{s}-\xi}\right\} = \frac{1}{\sqrt{\pi t}}e^{-\frac{z^2}{2\kappa t}} + \frac{V}{2\sqrt{\kappa}}e^{\frac{Vz}{2\kappa}\frac{V^2}{4\kappa}t}\text{erfc}\left(\frac{z}{2\sqrt{\kappa t}}+\frac{k}{2\sqrt{\kappa}}\sqrt{t}\right) \tag{7.292}$$

$$\pounds^{-1}\left\{\frac{e^{-\frac{z}{\sqrt{\kappa}}\sqrt{s}}}{\sqrt{s}+\xi}\right\} = \frac{1}{\sqrt{\pi t}}e^{-\frac{z^2}{2\kappa t}} - \frac{V}{2\sqrt{\kappa}}e^{\frac{Vz}{2\kappa}\frac{V^2}{4\kappa}t}\text{erfc}\left(\frac{z}{2\sqrt{\kappa t}}+\frac{k}{2\sqrt{\kappa}}\sqrt{t}\right) \tag{7.293}$$

Letting $\sqrt{s}+\xi = \sqrt{p}$ and using the definition of inverse Laplace integral $\pounds^{-1}\left\{\frac{e^{-\frac{z}{\sqrt{\kappa}}\sqrt{s}}}{\sqrt{s}+\xi}\right\}$ become as follow:

$$\pounds^{-1}\left\{\frac{e^{-\frac{z}{\sqrt{\kappa}}\sqrt{s}}}{\sqrt{s}+\xi}\right\} = \frac{1}{\sqrt{\pi t}}e^{-\frac{z^2}{2\kappa t}}e^{\frac{V^2}{4\kappa}t}\text{erfc}\left(\frac{z\sqrt{t}}{2\sqrt{\kappa}}+\frac{z}{2\sqrt{\kappa t}}\right)\left(1+\frac{Vz}{2\kappa}+\frac{V^2t}{2\kappa}\right) - \frac{V\sqrt{t}}{2\sqrt{\pi\kappa}}e^{-\frac{z^2}{4\kappa t}} \tag{7.294}$$

Using Eqs. 7.292–7.294, and after simplification, the Laplace inversion of $H_3(s)$ becomes

$$\pounds^{-1}\{H_3(s)\} = \frac{1}{16\kappa V}\left[e^{\frac{Vz}{\kappa}}\text{erfc}\left(\frac{z-Vt}{2\sqrt{\kappa t}}\right) - \left(1+\frac{Vz}{\kappa}+\frac{V^2t}{\kappa}\right)\text{erfc}\left(\frac{z+Vt}{2\sqrt{\kappa t}}\right)\right] + \frac{\sqrt{t}}{8\kappa\sqrt{\pi\kappa}}e^{-\frac{(z+Vtz)^2}{4\kappa t}} \tag{7.295}$$

Let $w_2^2 = \frac{V^2}{4\kappa}$, $k_1 = \frac{z}{\sqrt{\kappa}}$, $w_3^2 = \frac{V^2}{4\kappa} - V_1$ and $w_4 = \frac{V}{2\sqrt{\kappa}}$, then $H_4(s)$ becomes

$$H_4(s) = \frac{1}{\left((\sqrt{s})^2 - w_2^2\right)\left((\sqrt{s})^2 - w_3^2\right)((\sqrt{s}) - w_4)}\frac{1}{2\sqrt{\kappa}}e^{-(k_1\sqrt{s})} \tag{7.296}$$

and after using partial fraction expansion, $H_4(s)$ becomes

$$H_4(s) = \left[\begin{array}{l} \dfrac{D_1}{(\sqrt{s}-w_2)} + \dfrac{D_2}{(\sqrt{s}+w_2)} \dfrac{D_3}{(\sqrt{s}-w_3)} \dfrac{D_4}{(\sqrt{s}-w_3)} \\ + \dfrac{D_5}{(\sqrt{s}-w_4)} \end{array} \right] \frac{1}{2\sqrt{\kappa}} e^{-(k_1\sqrt{s})} \tag{7.297}$$

where

$$D_1 = \frac{1}{2w_2\left(w_2^2 - w_3^2\right)(w_2 + w_4)} \tag{7.298}$$

$$D_2 = \frac{1}{2w_2\left(w_2^2 - w_3^2\right)(-w_2 + w_4)} \tag{7.299}$$

$$D_3 = \frac{1}{2w_3\left(w_3^2 - w_2^2\right)(w_3 + w_4)} \tag{7.300}$$

$$D_4 = \frac{1}{2w_3\left(w_3^2 - w_2^2\right)(-w_2 + w_4)} \tag{7.301}$$

$$D_5 = \frac{1}{\left(w_4^2 - w_2^2\right)\left(w_4^2 + w_3^2\right)} \tag{7.302}$$

After substitution of Eqs. 7.298 through 7.302 into Eq. 7.297, then the Laplace inversion $\pounds^{-1}\{H_4(s)\}$ can be written as:

$$\pounds^{-1}\{H_4(s)\} = \frac{1}{2\sqrt{\kappa}} \left\{ \begin{array}{l} D_1\left[\left(\dfrac{1}{\sqrt{\pi t}}\right)e^{-\frac{k_1^2}{4t}} + w_2 e^{-w_2 k_1} e^{w_2^2 t_1} \operatorname{erfc}\left(-w_2\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)\right] \\ +D_2\left[\left(\dfrac{1}{\sqrt{\pi t}}\right)e^{-\frac{k_1^2}{4t}} - w_2 e^{-w_2 k_1} e^{w_2^2 t_1} \operatorname{erfc}\left(w_2\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)\right] \\ +D_3\left[\left(\dfrac{1}{\sqrt{\pi t}}\right)e^{-\frac{k_1^2}{4t}} + w_3 e^{-w_3 k_1} e^{w_3^2 t_1} \operatorname{erfc}\left(-w_3\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)\right] \\ +D_4\left[\left(\dfrac{1}{\sqrt{\pi t}}\right)e^{-\frac{k_1^2}{4t}} - w_3 e^{-w_3 k_1} e^{w_3^2 t_1} \operatorname{erfc}\left(w_3\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)\right] \\ +D_5\left[\left(\dfrac{1}{\sqrt{\pi t}}\right)e^{-\frac{k_1^2}{4t}} - w_4 e^{-w_4 k_1} e^{w_4^2 t_1} \operatorname{erfc}\left(w_4\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)\right] \end{array} \right\} \tag{7.303}$$

However, it is known from Eq. 7.275 that $T(z,t)$ can be written as:

$$\begin{aligned} T(z,t) = \pounds^{-1}\{\overline{T}(z,s)\} &= -2\kappa\mu\varpi_1 e^{-\frac{V}{2\kappa}\left[z+\frac{V}{2\kappa}t\right]}\pounds^{-1}\{H_4(s)\} \\ &\quad + \varpi_1 e^{-\mu z}\pounds^{-1}\{H_2(s)\} \\ &\quad - 8\kappa^2\frac{\rho VL}{k}\pounds^{-1}\{H_3(s)\} \end{aligned} \tag{7.304}$$

Note that

$$\pounds^{-1}\{H_2(s)\} = \frac{1}{V_1-\beta}\left(e^{-\beta t} - e^{-V_1 t}\right) \tag{7.305}$$

Substituting Eqs. 7.298–7.302, into Eq. 7.303, we get

$$\begin{aligned} T(z,t) = \frac{\kappa\mu\varpi_1}{\sqrt{\kappa}}e^{-\frac{V}{2t}\left(z+\frac{V}{2}t\right)} &\left\{ \begin{array}{l} \frac{1}{\sqrt{\pi t}}e^{-\frac{k_1^2}{4t}}[D_1+D_2+D_3+D_4+D_5] \\ \quad D_1\left[w_2 e^{-w_2 k_1}e^{w_2^2 t}\operatorname{erfc}\left(-w_2\sqrt{t}+\frac{k_1}{2\sqrt{t}}\right)\right] \\ +D_2\left[-w_2 e^{-w_2 k_1}e^{w_2^2 t}\operatorname{erfc}\left(w_2\sqrt{t}+\frac{k_1}{2\sqrt{t}}\right)\right] \\ +D_3\left[w_3 e^{-w_3 k_1}e^{w_3^2 t}\operatorname{erfc}\left(-w_3\sqrt{t}+\frac{k_1}{2\sqrt{t}}\right)\right] \\ +D_4\left[-w_3 e^{-w_3 k_1}e^{w_3^2 t}\operatorname{erfc}\left(w_3\sqrt{t}+\frac{k_1}{2\sqrt{t}}\right)\right] \\ +D_5\left[-w_4 e^{-w_4 k_1}e^{w_4^2 t}\operatorname{erfc}\left(w_4\sqrt{t}+\frac{k_1}{2\sqrt{t}}\right)\right] \\ +\frac{\varpi_1}{V_1-\beta}e^{-\mu z}\left[e^{-\beta t}-e^{-V_1 t}\right] \end{array} \right\} \\ &\quad -\frac{\alpha\kappa L}{2k}e^{-\frac{V_2}{\kappa}}\operatorname{erfc}\left(\frac{z-Vt}{2\sqrt{\kappa t}}\right) \\ &\quad -\left[\left(1+\frac{Vz}{\kappa}+\frac{V^2 t}{\kappa}\right)\operatorname{erfc}\left(\frac{z+Vt}{2\sqrt{\kappa t}}\right)\right] \\ &\quad -\frac{\rho VL}{k}\frac{\sqrt{\kappa t}}{\sqrt{\pi}}e^{-\frac{(z+Vt)^2}{4\kappa t}} \end{aligned} \tag{7.306}$$

Knowing that $D_1 + D_2 + D_3 + D_4 + D_5 = 0$ Eq. 7.306 becomes:

$$T(z,t) = \sqrt{\kappa}\mu\varpi_1 e^{-\frac{V}{2\kappa}\left(z+\frac{V}{2}t\right)} \times$$

$$\left\{ \begin{array}{l} \dfrac{e^{w_2^2 t}}{2\left(w_2^2 - w_3^2\right)} \left[e^{-w_2 k_1} \dfrac{\operatorname{erfc}\left(w_2\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)}{w_2 + w_4} + e^{-w_2 k_1} \dfrac{\operatorname{erfc}\left(w_2\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)}{w_4 + \mathrm{w}_2} \right] \\ + \dfrac{e^{w_3^2 t}}{2\left(w_3^2 - w_2^2\right)} \left[e^{-w_3 k_1} \dfrac{\operatorname{erfc}\left(w_3\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)}{w_3 + \mathrm{w}_4} + e^{-w_3 k_1} \dfrac{\operatorname{erfc}\left(w_3\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right)}{w_4 - w_3} \right] \\ + D_5 w_4 e^{\mathrm{w}_4^2 t} e^{\mathrm{w}_3^2 k_1} \operatorname{erfc}\left(w_4\sqrt{t} + \dfrac{k_1}{2\sqrt{t}}\right) \end{array} \right\}$$

$$+ \frac{\varpi_1}{V_1 - \beta} e^{-\mu z}\left[e^{-\beta t} - e^{-V_1 t}\right] - \frac{\alpha\kappa L}{2k} e^{\frac{Vz}{\kappa}} \operatorname{erfc}\left(\frac{z - Vt}{2\sqrt{\kappa t}}\right)$$

$$- \left[\left(1 + \frac{Vz}{\kappa} + \frac{V^2 t}{\kappa}\right) \operatorname{erfc}\left(\frac{z + Vt}{2\sqrt{\kappa t}}\right)\right]$$

$$- \frac{\rho V L}{k} \frac{\sqrt{\kappa t}}{\sqrt{\pi}} e^{-\frac{(z+Vt)^2}{4\kappa t}} \tag{7.307}$$

In order to generate graphical representation of Eq. 7.307, let us define the following variables, which will be later used in MathCAD program.

$$i = 0 \ldots N \ldots\ldots z_i = \frac{h_t i}{N} \tag{7.308}$$

$$j = 1 \ldots L \ldots\ldots t_j = \frac{\tau_i j}{L} \tag{7.309}$$

$$\tau_j = \kappa\mu^2 t \tag{7.310}$$

$$V^0 = \frac{V}{\kappa\mu} \quad \text{or} \quad V^0 = 1 + \frac{V_1}{\kappa\mu^2} \tag{7.311}$$

$$\beta^0 = \frac{1}{\kappa\mu^2}\beta \tag{7.312}$$

$$\mathrm{w}_2 = \frac{V^0}{4\kappa} - \beta^0 \tag{7.313}$$

$$\mathrm{w}_3 = \left(\frac{V^0 - 2}{2}\right)\kappa\mu^2 \tag{7.314}$$

$$\mathrm{w}_4 = \frac{V^0}{2}\sqrt{\kappa\mu} \tag{7.315}$$

$$D_5 = \frac{1}{\beta^0\left(V^0 - 1\right)\left(\kappa\mu^2\right)^2} \quad \text{or} \quad D_5 = \frac{1}{\beta V_1} \tag{7.316}$$

One can use MathCAD program to perform calculations, so our function $T(z,t)$ can be formulated T_{ij} and substituting above parameters we can obtain.

$$
\begin{aligned}
T_{i,j}(z,t) = {} & \frac{-I_{\max}}{k\mu}\exp\left[\frac{-V^{0^2}}{2}\left(z_d - \frac{V^0}{2}\tau_t\right)\right] \\
& \times \left\{\frac{\exp\left[\frac{V^{0^2}}{4}\left(V^0-1\right)\right]t_j}{\left[V^0-\left(1+\beta^0\right]\right.}\right\} \\
& \left[\frac{\exp\left[\left[\frac{V^{0^2}}{4}-\left(V^0-1\right)\right]t_j\right]}{\left[V^0-\left(1+\beta^0\right)\right]}\left[\begin{array}{l}
\exp\left[-\left(\sqrt{\frac{V^0}{4}}-\beta\right)\right]z_i\frac{1-\operatorname{erf}\left[-\sqrt{\left(\frac{V^{0^2}}{4}-\beta^0\right)}\right]t_j}{\sqrt{V^{0^2}-\beta^0}+V^0}+\frac{z_i}{2\sqrt{t_j}} \\
+\exp\left[\left(\sqrt{\frac{V^{0^2}}{4}}-\beta^0\right)z_i\right]\frac{1-\operatorname{erf}\left[\sqrt{\left(\frac{V^{0^2}}{4}-\beta^0\right)t_j}+\frac{z_i}{2\sqrt{t_j}}\right]}{\left(V^0-\sqrt{V^{0^2}-\beta^0}\right)}
\end{array}\right] \\
& +\frac{\exp\left[\left[\frac{V^{0^2}}{4}-\left(V^0-1\right)\right]t_j\right]}{\left[\beta^0-\left(V^0-1\right)\right]}\left[\begin{array}{l}
\exp\left[\sqrt{\frac{V^{0^2}}{4}}-\left(V^0-1\right)\right]\frac{1-\operatorname{erf}\left[-\sqrt{\left[\frac{V^{0^2}}{4}-\left(V^0-1\right)\right]t_j+\frac{z_i}{2\sqrt{t_j}}}\right]}{\sqrt{V^{0^2}-4\left(V^0-1\right)}+V^0} \\
+\exp\left[\sqrt{\frac{V^{0^2}}{4}}-\left(V^0-1\right)\right]\frac{1-\operatorname{erf}\left[\sqrt{\left[\frac{V^{0^2}}{4}-\left(V^0-1\right)\right]t_j+\frac{z_i}{2\sqrt{t_j}}}\right]}{-\sqrt{V^{0^2}-4\left(V^0-1\right)}+V^0} \\
+\frac{-1}{2\beta^0\left(V^0-1\right)}\exp\left[\frac{V^{0^2}}{4}t_j\right]\exp\left[\frac{V^0}{2}(z_i)\right]\left(1-\operatorname{erf}\left(\frac{V^0}{2}t_j+\frac{z_i}{2\sqrt{t_j}}\right)\right)
\end{array}\right] \\
& +\frac{\kappa\rho L}{2k}\left[\begin{array}{l}
e^{V^0 z_i}\left(1-\operatorname{erf}\left(\frac{z_i-V^0t_j}{2\sqrt{t_j}}\right)\right)-\left(1+V^0z_1+V^{0^2}t_j\right)\left(1-\operatorname{erf}\left(\frac{z_i+V^0t_j}{2\sqrt{t_j}}\right)\right) \\
+\frac{2V^0t_j}{\sqrt{\pi}}\exp\left[-\left(\frac{V^{0^2}}{4\tau_j}-\frac{V^0z_i}{2}+\frac{V^{0^2}t_j}{4}\right)\right]
\end{array}\right] \\
& +\left[\frac{I_{\max}}{k\mu}\frac{1}{\left(V^0-1\right)-\beta^0}e^{-z_i}\left[\exp\left(e^{-\beta^0t_j}-e^{-\left(V^0-1\right)t_j}\right)\right]\right]
\end{aligned}
\tag{7.317}
$$

$$
E_{1_{i,j}} = \sqrt{\left(\frac{V^{0^2}}{4}-\beta^0\right)t_j}+\frac{z_i}{2\sqrt{t_j}} \tag{7.318}
$$

$$
E_{2_{i,j}} = \sqrt{\left[\frac{V^{0^2}}{4}-\left(V^0-1\right)\right]t_j+\frac{z_i}{2\sqrt{t_j}}} \tag{7.319}
$$

$$E_{3_{i,j}} = \sqrt{V^{0^2} - 4\left(V^0 - 1\right)} \tag{7.320}$$

$$E_{4_{i,j}} = e^{-\left[\sqrt{\frac{V^{0^2}}{4} - \left(V^0 - 1\right)}\right] z_1} \tag{7.321}$$

$$E_{5_{i,j}} = e^{-\left[\sqrt{\left(\frac{V^{0^2}}{4} - \beta^0\right)}\right] z_1} \tag{7.322}$$

$$E_{6_{i,j}} = \frac{V^0}{2} t_j + \frac{z_i}{2\sqrt{t_j}} \tag{7.323}$$

$$E_{7_{i,j}} = \frac{z_i + V^0 t_j}{2\sqrt{t_j}} \tag{7.324}$$

$$E_{8_{i,j}} = \frac{V^{0^2}}{4} - \left(V^0 - 1\right) \tag{7.325}$$

$$T(z,t) = \frac{-I_{\max}}{k\mu} \exp\left[\frac{-V^{0^2}}{2}\left(z_d - \frac{V^0}{2}\tau_t\right)\right] \times$$

$$\left[\begin{array}{l}
\dfrac{\exp\left(E_{8_{d,t}}\right)\tau_t}{\left[V^0 - \left(1+\beta^0\right)\right]}\left[E_{5_{d,t}} \dfrac{1 - \operatorname{erf}\left(-\operatorname{Re}\left(E_{1_{d,t}}\right)\right)}{\sqrt{V^{0^2} - \beta^0} + V^0} + E_{5_{d,j}} \dfrac{1 - \operatorname{erf}\left(-\operatorname{Re}\left(E_{1_{d,t}}\right)\right)}{V^0 - \sqrt{V^{0^2} - \beta^0}}\right] \\
+ \dfrac{e^{E_{8_{d,s}}\tau_t}}{\left[\beta^0 - \left(V^0 - 1\right)\right]}\left(E_{4_{d,t}} \dfrac{1 - \operatorname{erf}\left(-E_{2_{d,t}}\right)}{E_{3_{d,t}} + V^0} + E_{4_{d,t}} \dfrac{1 - \operatorname{erf}\left(-E_{2_{d,t}}\right)}{-E_{3_{d,t}} + V^0}\right) \\
- \dfrac{1}{2\beta^0\left(V^0 - 1\right)} \exp\left[\dfrac{V^{0^2}}{4}\tau_t\right] \exp\left[\dfrac{V^0}{2} z_d\right] \left(1 - \operatorname{erf}\left(E_{6_{d,t}}\right)\right) \\
- \dfrac{\kappa\rho L}{2k}\left[\begin{array}{l} e^{-V^0 z_d}\left(1 - \operatorname{erf}\left(\dfrac{z_d - V^0\tau_t}{2\sqrt{\tau_t}}\right)\right) \\ -\left(1 - V^0 z_d + V^{0^2}\tau_t\right)\left(1 - \operatorname{erf}\left(E_{7_{d,t}}\right)\right) \end{array}\right] \\
+ \dfrac{2V^0\tau_t}{\sqrt{\pi}} \exp\left[-\left(\dfrac{V^{0^2}}{4\tau_t} - \dfrac{V^0 z_d}{2} + \dfrac{V^{0^2}\tau_t}{4}\right)\right] \\
+ \left[\dfrac{I_0}{k\mu\left(V^0 - 1\right) - \beta^0} \exp\left[-z_d\left(e^{-\beta^0\tau_t} - e^{-\left(V^0 - 1\right)\tau_t}\right)\right]\right]
\end{array}\right] \tag{7.326}$$

Temperature distribution with respect to depth of the workpiece and time is graphically shown in Fig. 7.56, where recession velocity of vaporization surface was estimated to be about 33 m/s.

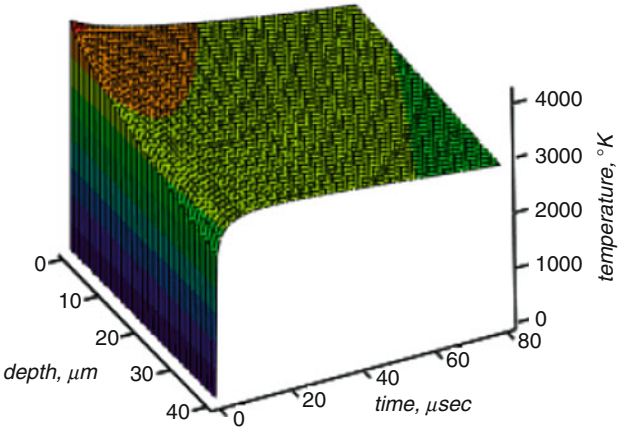


Fig. 7.56 Temperature distribution of the workpiece with consideration of transfer evaporative case calculated using Eq. 7.326

A closed form solution illustrated in Fig. 7.56 is limited to the surface ablation only and does not consider the plasma formation and liquid expulsion from irradiated workpiece.

The temperature gradient attains relatively lower values in the surface vicinity of the substrate material than that corresponding to some depth below the surface. In this case, internal energy gain in this region due to absorption of laser beam energy becomes more important as compared to diffusional heat transfer from the surface vicinity to the bulk of the substrate material due to temperature gradient. As the heating period progresses, the point of minimum temperature gradient moves towards the solid bulk of the substrate material.

In the region beyond the point of minimum temperature gradient, the diffusional energy transport plays an important role inside the substrate material.

The large change in the power intensity occurs with increasing β^0; however, the magnitude of temperatures in the surface region does not alter considerably. This is because of the rate of evaporation, which increases with increasing power intensity. Consequently, the convective boundary condition at the surface suppresses the temperature rise in the surface vicinity at high-power intensities

7.9.1.7 Analytical Solution of Heat Transfer Equation with Kinetic Theory Approach

This section examines pulsed laser heating process by considering both Fourier conduction and electron–phonon kinetic theory approaches, which is described in Sect. 7.10 of this chapter later on. A one-dimensional kinetic theory approach is presented. More specifically, this section is focused on development of equations governing an interaction of a Gaussian laser beam with materials, where analytical

solution of these equations based on both the Fourier heat conduction theory and kinetic theory approach using a Laplace integral transformed method is utilized. A comparison of the analytical solution of the Fourier theory and closed form solution of the kinetic theory approach is introduced in the next section of this chapter (Sect. 7.11). The temperature distribution for the heated material predicted from the kinetic theory is compared with the Fourier theory findings.

Most theoretical work relating to laser beam interaction with materials is based on the solution of the classical heat conduction equation derived from Fourier theory. It has been shown that the Fourier theory of heat conduction is not fully applicable to short pulsed laser heating due to the assumptions made in the theory.

These assumptions include:

1. Matter is assumed as continuous and homogeneous, and that;
2. The heat flux across any plane is a function of only the temperature gradient at that plane.

The first assumption is not valid for distances less than inter-atomic spacing and the second is only true if all the energy crossing the isothermal plane is accounted for.

Conduction in metals occurs due to subsequent collisions between excited electrons and lattice site atoms. The electron motion in the substrate is random, which means that electrons move from the surface to the bulk as well as from the bulk to the surface. Moreover, the amount of energy transferred to lattice site atoms depends on the electron energy distribution in a particular region. Harrington (1967) showed, based on kinetic theory considerations, that electrons within five times of the electron mean free path contribute 98.5 % of the total energy transported provided that $\partial T/\partial z$ is constant over this distance. Consequently, attempting to generate the Fourier equation, the application of $q = -k(\partial T/\partial z)$ is limited to planes in excess of ten times the electron mean free path apart.

In addition, the absorption depth of the metals is of the order of ten electron mean free paths; therefore, the gradient $\partial T/\partial z$ is not uniform over the spatial increment $(\Delta z \geq 10\lambda)$ and λ being the electron mean free path. In this case, the higher order gradients $\left(\partial^3 T/\partial z^3\right)$, which are neglected in the Fourier heating model, become important and the validity of the Fourier heating model comes into question. Therefore, it becomes necessary to examine the laser-induced conduction heating on a microscopic scale.

The applicability of the Fourier equation in laser heating is limited to the cases in which low power laser intensities are employed [56].

This is due to the following facts:

1. In the analysis of the Fourier heat conduction model, the heat flux through a given plane is considered as being a function of the spatial temperature gradient at that plane. This depends upon the assumption that the temperature gradient remains almost constant between two successive and closely spaced planes. The distance between these planes is finite, and therefore, error occurs when

high-order terms, which are neglected, become important at high-power laser intensities.
2. The heat flux through a given plane depends on the electron energy distribution through the material, therefore, the material cannot be considered as a homogeneous continuum.

Consequently, a new model may be required to be developed for heating mechanism, which is appropriate to high-power laser heating process. A model considering a kinetic theory approach describing the transport of energy by electrons within the electron mean free path may be suitable in this case. The basis of this model was introduced by Yilbas [57] for one-dimensional heating. It was shown that the predictions made from the new model agree well with the experimental findings. He adopted the electron motion in metals to formulate the laser pulse heating process. The heating process was established based on a kinetic energy transfer mechanism, which occurred during the collisions between excited electrons and lattice site atoms. In this case, the excess electron energy was transferred to lattice site atoms resulting in increased amplitude of lattice site vibrations during the collision process.

In this section, comparisons of the electron kinetic theory approach are made with Fourier theory models for a pulse laser heating process. The temperature field due to each model is predicted for step intensity as well as exponentially decaying intensity pulses. The study is extended to include the analytical solution to the electron kinetic theory approach for Gaussian intensity pulses. Electron kinetic theory approach is based on electron and phonon movements.

7.9.2 Melting Process

Consider first the case of a semi-infinite slab as before, melted, with instantaneous melt removal, as indicated in Fig. 7.57.

We solve this by considering ourselves moving along with interface. But to do this we have to reconsider our heat-flow equation, which is:

$$\frac{\partial^2 T(z,t)}{\partial z^2} - \frac{2}{\kappa}\frac{\partial T(z,t)}{\partial t} = 0 \tag{7.327}$$

Recall that this was derived by noting the rate at which heat accumulated in an elemental volume:

$$\rho c \frac{\partial T(z,t)}{\partial t} + \frac{\partial}{\partial z}\left(-K\frac{\partial T(z,t)}{\partial z}\right) = 0 \tag{7.328}$$

If the medium is moving, then an additional amount of heat ρcT is flowing in at rate V, therefore

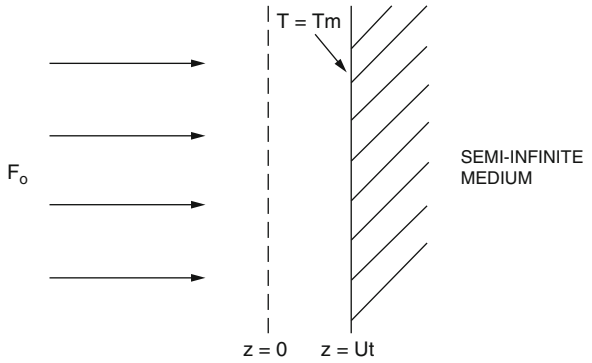


Fig. 7.57 Irradiation of a semi-infinite slab with instantaneous melt removal

$$-K\frac{\partial T}{\partial z} \quad \text{becomes} \quad -K\frac{\partial T}{\partial z} + V\rho cT$$

As result of this, the Eq. 7.329a turns out to be:

$$\rho c\frac{\partial T(z,t)}{\partial t} + \frac{\partial}{\partial z}\left(-K\frac{\partial T(z,t)}{\partial z} + V\rho cT(z,t)\right) = 0 \tag{7.329a}$$

or for simplicity, we write as

$$\rho c\frac{\partial T}{\partial t} + \frac{\partial}{\partial z}\left(-K\frac{\partial T}{\partial z} + V\rho cT\right) = 0 \tag{7.329b}$$

This equation is valid, providing that we are not generating any heat in the solid. Assuming that $\kappa = K/\rho c$ as before, then Eq. 7.329b yields to:

$$\frac{\partial T}{\partial t} - \kappa\frac{\partial^2 T}{\partial z^2} + V\frac{\partial T}{\partial z} = 0 \tag{7.330}$$

Therefore, we can find at least a steady-state solution to the problem by basing our coordinate system at the interface and letting the material flow in at some rate to be determined. Call the rate $-U$, where U is a positive number. The solution will be

$$T = T_m e^{-Uz'/\kappa} \tag{7.331}$$

where κ is diffusivity of the solid, T_m is the melting temperature, and z' is the distance from melting front. That is, this is solution to:

$$\frac{\partial T(z',t)}{\partial t} - \kappa\frac{\partial^2 T(z',t)}{\partial z'^2} - U\frac{\partial T(z',t)}{\partial z'} = 0 \tag{7.332}$$

as we can verify, First, $\partial T/\partial t = 0$ because this is the steady-state profile, or

$$\kappa \frac{\partial^2 T(z',t)}{\partial z'^2} + U \frac{\partial T(z',t)}{\partial z'} = 0 \tag{7.333}$$

of course $T = T_m$ at $z' = 0$ and $T(z',t) \to 0$ as $z' \to 0$. It is trivial to show, by differentiation, that Eq. 7.330 is the solution to Eq. 7.332.

To calculate U, we use energy balance. F_0 must raise the material to T_m, its melting point and note that T_m °C above ambient or environment, and then melt it. Thus, in time Δt the energy put into thickness $\Delta z^{'}$ (where $U = \Delta z^{'}/\Delta t$) must be given by:

$$\frac{F_0 \Delta t}{\Delta z'} = L\rho + cT_m\rho \tag{7.334a}$$

or

$$F_0 = \rho[L + cT_m]U \tag{7.334b}$$

Thus, we can write, for the steady-state temperature profile,

$$T(z',t) = T_m \exp\left\{-\frac{z'}{\kappa}\left[\frac{F_0}{\rho(L + cT_m)}\right]\right\} \tag{7.335}$$

Definition of Sublimation Sublimation is a physical process in which a solid directly converts into a gaseous (vapor) state without going through a liquid state. The latent heat of sublimation at a particular temperature is the amount of heat required to convert a unit mass of solid into gas. For example, when ice sublimates into vapor, the amount of heat required at 0 °C is estimated equal to 2,838 kJ/kg, which is the latent heat of sublimation of ice at 0 °C. In the crystal growth of ice and snow in atmosphere, this process plays a dominant role. See Fig. (d) below

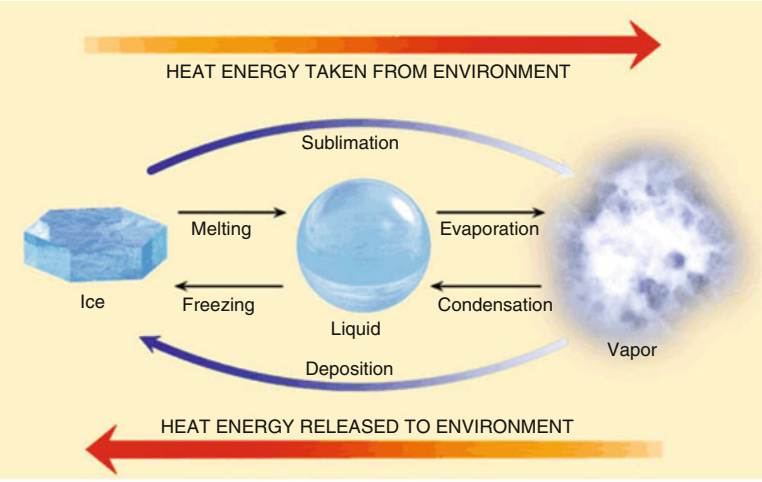


Fig. d Depiction of Latent Heat

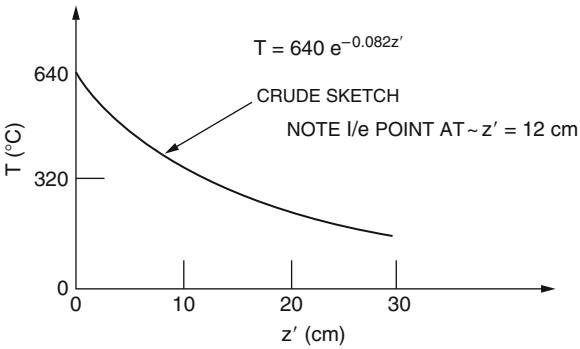


Fig. 7.58 Temperature profile in aluminum with instantaneous melt removal

This process is opposite to the process of deposition. Also bear in your mind that, when a substance changes from one state to another, by definition, latent heat is added or released in the process.

where $F_0/[\rho(L + cT_m)]$ is the velocity of the melting front. This solution also would be appropriate to sublimation, where L is then the latent heat of sublimation and T_m the sublimation temperature. Note: that this is the semi-infinite-slab approximation and cannot be used to estimate the time to penetrate a slab of given thickness.

Let us consider aluminum, with $F_0 = 200\,\mathrm{W/cm^2}$. Taking a more accurate value of the melting point than in earlier examples, $T_m = 640\,^\circ\mathrm{C}$ (the actual melting point of 660 °C minus room temperature of 20 °C). If we put in the other values of the parameters, T falls off as shown in Fig. 7.58.

As another illustration, consider Plexiglas. Plexiglas is rapidly eroded by 10.6 μm laser radiation, by a process that is essentially sublimation. Since it couples extremely well ($A \approx 1.0$) and has a very low thermal diffusivity $\left(\kappa \approx 10^{-4}\,\mathrm{cm/s^2}\right)$, it can be used to make "burn patterns" of the beam. That is, the depth of the erosion at a given point is linearly proportional to the energy density incident at that point. We can understand this by applying heat-flow concept. Let us apply Eq. 7.330 and interpret L as the heat of erosion. Since $c \approx \mathrm{J/g} - {}^\circ\mathrm{C}$ and $T_m = 200\,^\circ\mathrm{C}$ and $L \approx 1000\,\mathrm{J/g}$, for a crude calculation we can ignore cT_m. The density is about 1.1 g/cm^3. We then have:

$$T(z', t) \approx 200\ e^{-9F_0 z'} \tag{7.336}$$

Since $A \approx 1.0$, for a typical power density like 5 kW/cm^2 we have $F_0 = 5 \times 10^3$, so that

$$T \approx 200\ e^{-4.5\times 10^4 z'} \tag{7.337}$$

Hence, the temperature profile is confined to an extremely narrow region near the eroding surface. The rate of erosion is, by Eq. 7.334b with cT_m ignored as:

$$U \approx 4.5\,\mathrm{cm/s} \tag{7.338}$$

Now we can see why Plexiglas is useful for monitoring the beam profile, and why it can pick up fairly fine structure in the beam. After irradiation for a time t the penetration depth is $Ut = (F_0 t)/(\rho L)$, and this should be large compared to a thermal diffusion length $D = 2\sqrt{\kappa t}$ if the pattern is to reveal fine structure. Otherwise thermal diffusion would "wash out" the pattern by distributing the energy in a radial direction. Thus

$$2\sqrt{\kappa t} \ll \frac{F_0 t}{\rho L} \tag{7.339}$$

For the numbers we used above at $t = 1/4$ s, we have a depth of 1 cm. Thus

$$2\sqrt{10^{-4} \times \frac{1}{4}} = 0.1 \text{ mm} \ll 1 \text{ cm} \tag{7.340}$$

and we see the criterion is well satisfied.

Let us look at a more complete problem, namely melting by laser radiation of a slab of material. One basic problem is what happens to the melted material. (We ignore the vaporization question for now.) There are two eases which are fairly amenable to numerical solution. They are the "fully retained liquid" case, in which all the liquid is presumed to stay in place, and the "full ablation" case, in which the melt is presumed to disappear magically as soon as it forms. The latter case might correspond to the presence of a heavy wind-stream which blows away the melt.

Looking first at the full ablation case, we have the situation shown in Fig. 7.59. T_2 is the temperature in the solid, and the front surface at $z = 0$ first warms up to the melting point T_m (above ambient temperature) and then begins to move to the right. S denotes its position as a function of time. When $S = \ell$, the process is over, and we call this time t_f. We denote by t_m the time at which the front surface begins to melt. The field equation is:

$$\frac{\partial^2 T_2}{\partial z^2} - \frac{1}{\kappa_2}\frac{\partial T_2}{\partial t} = 0 \quad \text{for} \quad 0 \le t \le t_f \tag{7.341}$$

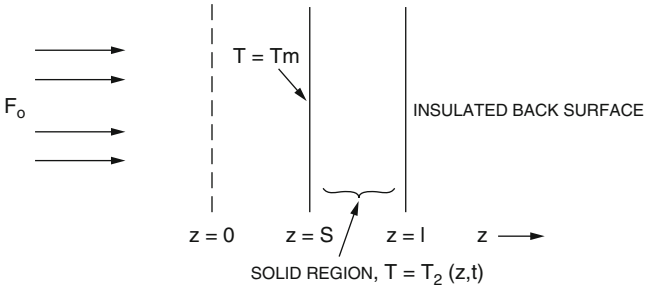


Fig. 7.59 Fully ablated case

The boundary conditions are

$$K_2 \frac{\partial T_2}{\partial z}\bigg|_{z=0} = -F_0 \quad \text{for} \quad 0 \le t \le t_m \tag{7.342}$$

$$K_2 \frac{\partial T_2}{\partial z}\bigg|_{z=S} = -F_0 + \rho L \frac{dS}{dt} \quad \text{for} \quad t_m \le t \le t_f \tag{7.343}$$

$$K_2 \frac{\partial T_2}{\partial z}\bigg|_{z=\ell} = 0 \quad \text{for} \quad 0 \le t \le t_f \tag{7.344}$$

The boundary conditions are:

$$\begin{aligned} T_2(z,0) &= T_{20} \\ S|_{t \le t_m} &= 0 \end{aligned} \tag{7.345}$$

The above boundary conditions are nonlinear, and a solution in analytical form is very difficult. This is due to the presence of the moving boundary and appears in the second boundary condition, which states that the boundary moves at a rate dS/dt determined by a balance between the heat of melting L, the heat input F_0, and the heat flow by thermal conduction.

One relationship must hold for this problem; it follows from energy balance. The total energy put in per unit area is $F_0 t_f$, and, since the material is simply heated to T_m and melted, this energy goes solely to those processes. Thus

$$F_0 t_f = \rho(L + cT_m) \tag{7.346}$$

This is convenient, for one can check numerical solutions. More important, it gives a first-order estimate of the time needed to melt through materials by laser radiation.

Turning now to the fully retained liquid case, we have the following set of equations. The definitions are the same as above, except that subscript 1 now refers to the molten state, whereas subscript 2 is still the solid state (see Fig. 7.60). The field equations are

$$\frac{\partial^2 T_2}{\partial z^2} - \frac{1}{\kappa_2}\frac{\partial T_2}{\partial t} = 0 \quad \text{for} \quad 0 \le t \le t_f \quad \text{(solid)} \tag{7.347a}$$

$$\frac{\partial^2 T_1}{\partial z^2} - \frac{1}{\kappa_1}\frac{\partial T_1}{\partial t} = 0 \quad \text{for} \quad t_m \le t \le t_f \quad \text{(liquid)} \tag{7.347b}$$

The boundary conditions are:

$$K_2 \frac{\partial T_2}{\partial z}\bigg|_{z=0} = -F_0 \quad \text{f} \quad 0 \le t \le t_m \tag{7.348a}$$

$$K_1 \frac{\partial T_1}{\partial z}\bigg|_{z=0} = -F_0 \quad 0 \le t \le t_f \tag{7.348b}$$

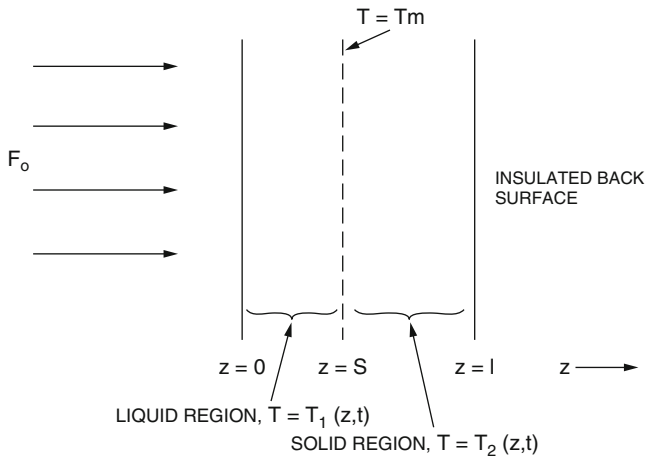


Fig. 7.60 Fully retained liquid case

$$K_1 \frac{\partial T_1}{\partial z}\bigg|_{z=S} - K_2 \frac{\partial T_2}{\partial z}\bigg|_{z=S} = -\rho L \frac{dS}{dt} \quad t_m \le t \le t_f \tag{7.348c}$$

$$K_2 \frac{\partial T_2}{\partial z}\bigg|_{z=\ell} = 0 \quad 0 \le t \le t_f \tag{7.348d}$$

$$T_2|_{z=S} = T_1|_{z=S} = T_m \quad t_m \le t \le t_f \tag{7.348e}$$

The initial conditions are:

$$T_2(z,0) = T_{20} \tag{7.349a}$$

$$S|_{t \le t_m} = 0 \tag{7.349b}$$

Rather than discuss this problem in detail, we pass on to the more practical, although more complex, case of vaporization. For the fully retained liquid case, suffice it to say that the retained liquid has a shielding effect, and this causes the time to reach melting at the back surface to be longer than in the ablated model. Some typical values of melt through time for 0.2-cm-thick material with $F_0 \approx 2$ kW/cm^2 are given below.

	Ablated (s)	Retained (s)
Aluminum	0.32	0.37
Stainless steel	4.0	4.5

7.9.3 Melting and Vaporization

We present here without derivation some results for the case of a slab of material, insulated on the surfaces, subjected to uniform and continuous irradiation [58]. These are one-dimensional calculations. It is assumed that the melt is fully retained until it reaches the vaporization temperature, where it disappears. Then we have the case illustrated in Fig. 7.61.

Here S_2 is the, position of the liquid–vapor interface and S_1 the position of the solid–liquid interface. This problem is been solved numerically at NRL [58], and we shall show sonic results. The assumptions are that in each phase the thermal properties are independent of temperature.

In the curves, the following definitions are used:

$$\alpha\hat{Q} = \frac{F_0\ell}{\rho L\kappa_{\text{solid}}} \tag{7.350}$$

$$\tau = \frac{\kappa_{\text{solid}}}{\ell^2} \tag{7.351}$$

$$\theta = (T - T_m)c_{\text{solid}}/L \tag{7.352}$$

where

$$L = \text{latent heat of fusion}$$

Subscription 0 is = ambient or environment temperature

$$\omega = L_v/L$$

with L_v is the latent heat of vaporization.

In these equation T is understood to be in degree Celsius and represents the actual temperature. Although the thermal conductivities of liquid and solid are allowed to differ, the specific heats are assumed to be the same.

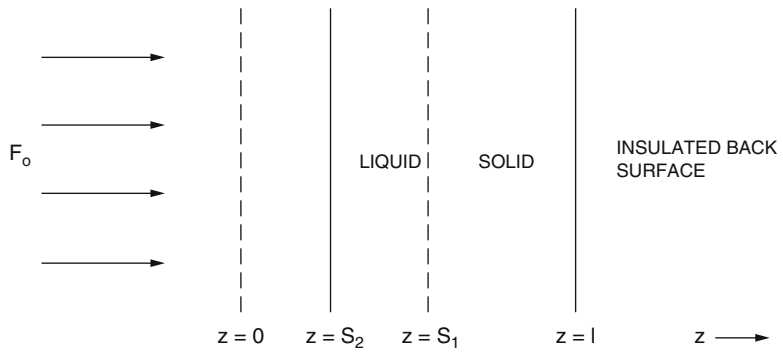


Fig. 7.61 Melting and vaporization, with fully retained liquid

Note that when $F_0 \to 0$ and $F_0 \to \infty$, we get certain easy 'limits. For $F_0 \to 0$ no vaporization can take place, the melting is small, and we approach toe fully ablated limit. On the other hand, as $F_0 \to \infty$ all the liquid should be vaporized by (f, the time the back surface melts.

So in this limit

$$F_0 t_f = \rho \ell \left[L + (T_m - T_0) c_{\text{solid}} + (T_{\text{v}} - T_m) c_{\text{liquid}} + L_v \right] \quad \text{for} \quad F_0 \to \infty \tag{7.353}$$

whereas

$$F_0 t_f = \rho \ell [L + (T_m - T_0) c_{\text{solid}}] \quad \text{for} \quad F_0 \to 0 \tag{7.354}$$

The limiting values are represented by asymptotes of the curves Figs. 7.62–7.67 indication Fig. 7.64 indicating by dashed lines on the plots of $\alpha \hat{Q}$ vs. τ. Note on these plots that the dashed lines are at 45°, or have a slope of −1. These are log-log plots, so the asymptotes can be described by

$$\log\left(\alpha \hat{Q}\right) = \log C - \log \tau_f \tag{7.355}$$

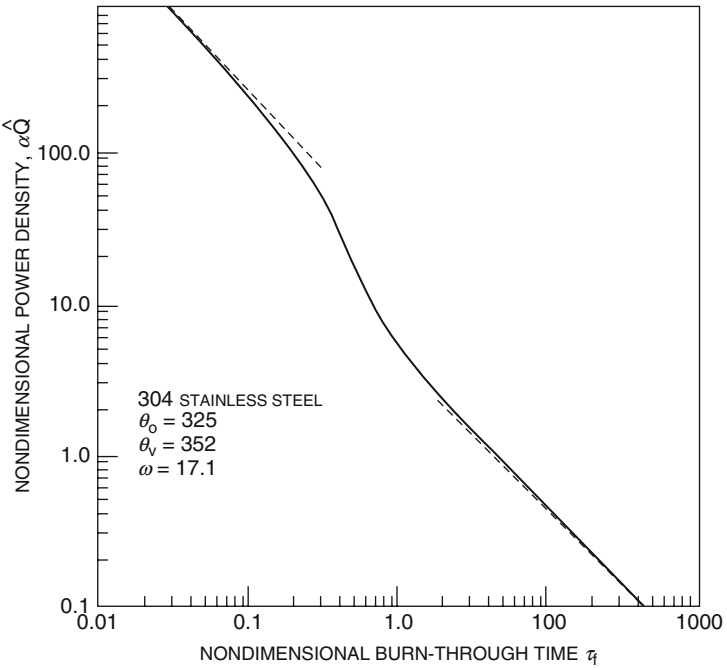


Fig. 7.62 Power density vs. burn-through time for stainless steel [58]

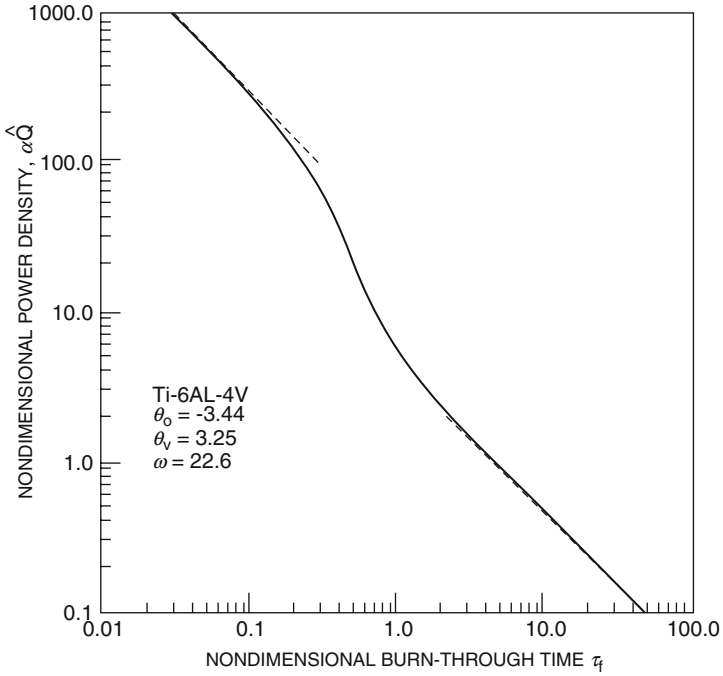


Fig. 7.63 Power density vs. burn-through time for titanium alloy [58]

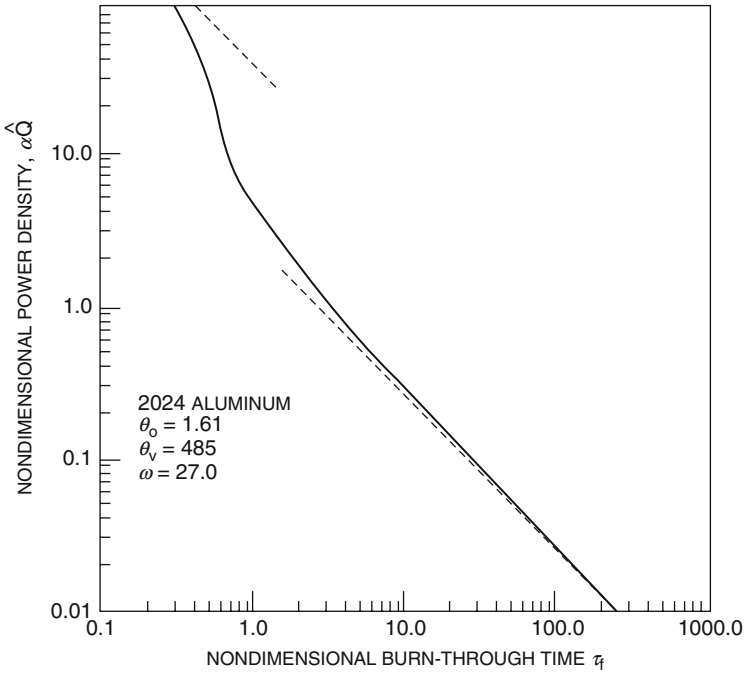


Fig. 7.64 Power density vs. burn-through time for aluminum alloy [58]

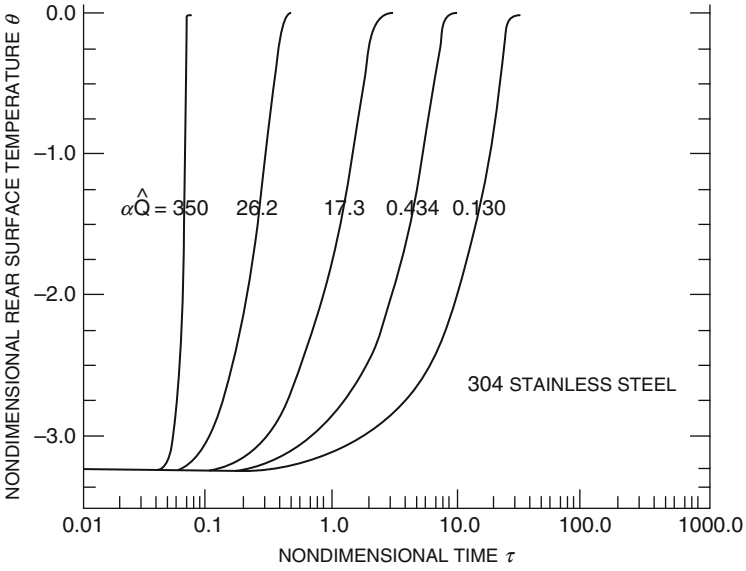


Fig. 7.65 Rear surface temperature rise for stainless steel [58]

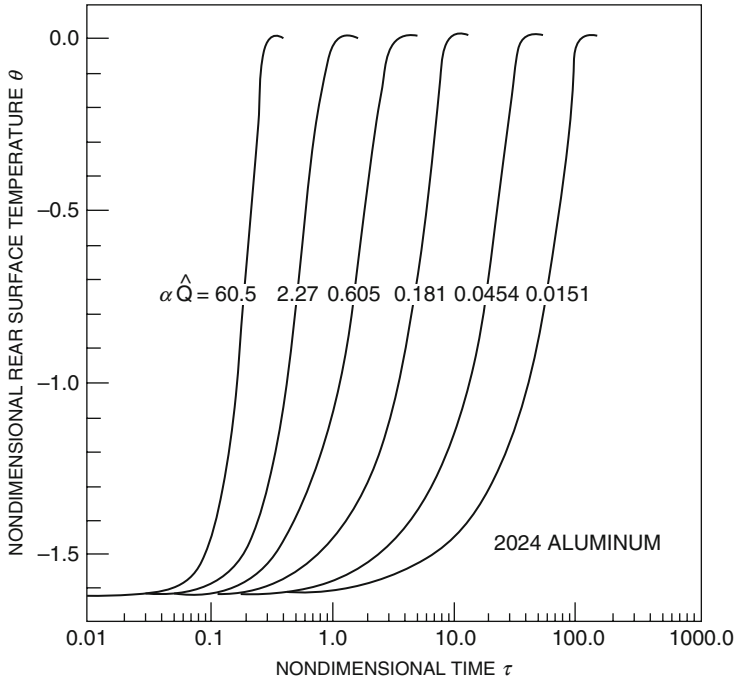


Fig. 7.66 Rear surface temperature rise for aluminum alloy [58]

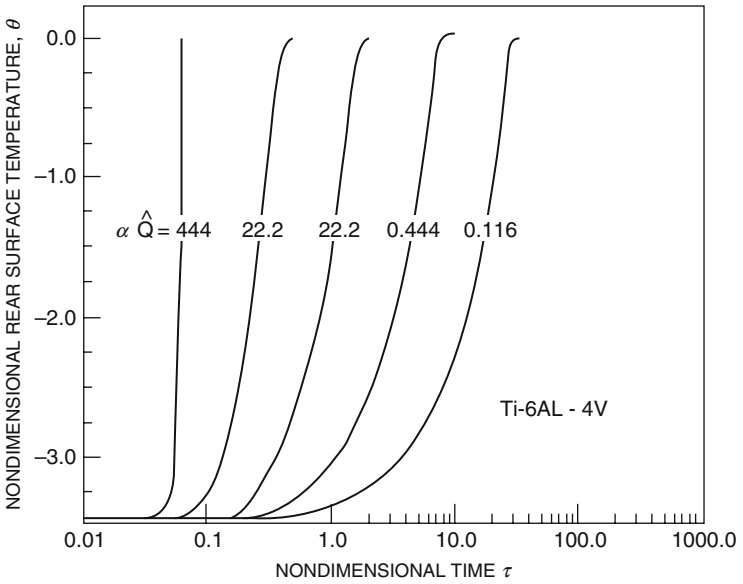


Fig. 7.67 Rear surface temperature rise for titanium alloy [58]

where $C \to C_\infty$ as $\alpha\hat{Q} \to \infty$, and $C \to C_0$ as $\alpha\hat{Q} \to 0$. If we take the antilog of Eq. 7.355, we get the following:

$$\alpha\hat{Q} = \frac{C}{\tau_f} = C(\tau_f)^{-1} \tag{7.356}$$

Substituting in the definition of $\alpha\hat{Q}$ Eq. 7.350 and τ Eq. 7.351, gives:

$$\frac{F_0\ell}{\rho L\kappa_{\text{solid}}} = C\frac{\ell^2}{\kappa_{\text{solid}}\tau_f} \tag{7.357a}$$

or

$$F_0\tau_f = \rho\ell LC \tag{7.357b}$$

By comparison with the $F_0 \to 0$ and $F_0 \to \infty$ limits above we can see that

$$C_0 = \frac{1}{L}[L + (T_m - T_0)c_{\text{solid}}] \tag{7.358}$$

and

$$C_\infty = \frac{1}{L}\left[L + (T_m - T_0)c_{\text{solid}} + (T_{\text{v}} - T_m)c_{\text{liquid}} + L_{\text{v}}\right] \tag{7.359}$$

The numerical values of the thermal parameters which were used in generating these solutions are included as a separate table (Table 7.4) in addition to the graphs (Figs. 7.62, 7.63, 7.64, 7.65, 7.66, and 7.67).

Table 7.4 Thermal properties of metals and alloys

Metal or alloy	T_m (°C)	T_v (°C)	L (cal/g)	L_v (cal/g)	ρ (g/cm^3)	K_S (cal/s-cm-°C)	K_Q (cal/s-cm-°C)	c_S (cal/g-°C)	c_Q (cal/g-°C)	k_S (cm^2/s)	k_Q (cm^2/s)
304 Stainless steel	1454	3000	65	1112	8.0	0.062	0.062	0.148	0.148	0.0523	0.0523
Ti (6Al, 4 V)	1649	3285	94	2124	4.47	0.0455	0.0455	0.199	0.199	0.051	0.051
2024 Alum.	638	2480	95	2565	2.77	0.414	0.207	0.249	0.249	0.6	0.3

Provided one knows F_0, these solutions are reasonable estimates for the time to penetrate a metal specimen with a laser beam. In application, however, one must consider the actual size of the beam. These solutions will be useful for effects in the center of the beam if the diffusion length is small compared to the beam radius, or if for times up to and including the melt-through time t_f the beam radius $R > 2\sqrt{\kappa t_f}$ is as:

$$R > 2\sqrt{\kappa t_f} \tag{7.360}$$

In terms of the parameter τ_f this becomes, upon squaring both sides as:

$$\tau_f =< \frac{1}{4}\left(\frac{R}{\ell}\right)^2 \tag{7.361}$$

7.9.4 Electron–Phonon Analytical Solution

The kinetic theory approach essentially deals with the kinetic energy transfer mechanism that occurs when electrons and lattice atoms with different energies interact. In order to simplify the phenomenon, some useful assumptions are made. These include omission of thermionic emission, attainment of steady space charge, and mean free path of molecules being independent of temperature. A net flow of electrons occurs in the substrate by the presence of an electron source at infinity, and the electron gives fraction of its excess energy to the lattice site atoms during an electron–phonon collision and this fraction is assumed to be constant throughout the successive collisions.

When the electron absorbs the incident laser energy, some excess energy of the electron is transferred to the lattice site atoms during electron–phonon collision process. This energy manifests itself as an increase in the amplitude of the atomic vibration (phonon). As a result, neighboring atoms in the lattice are forced away to new equilibrium positions and absorb some of this extra energy in the process. A stage may be reached where eventually the lattice site atoms in the localized region around the original collision site are all in equilibrium and have increased their vibrational energies. It is this energy mechanism, which defines the conduction process in the solid substrate when subjected to a laser heating pulse. The amount of energy which electrons from section I transfer to lattice site atoms in the same section, Fig. 7.68, can be calculated as follows.

The number of electrons leaving section I, is: $N_z A_z \overline{V}_z dt$ where $A_z = dxdy$, and N_z is the number density of electrons which transfer energy to dz from $d\zeta$, and $\overline{V}_z$ is the average electron velocity entering the control volume in the z-axis across area A_z at time dt.

The energy of N atoms in one direction due to lattice vibration is: $E = Nk_B T$ which can be described as the phonon energy. The number of collisions, which

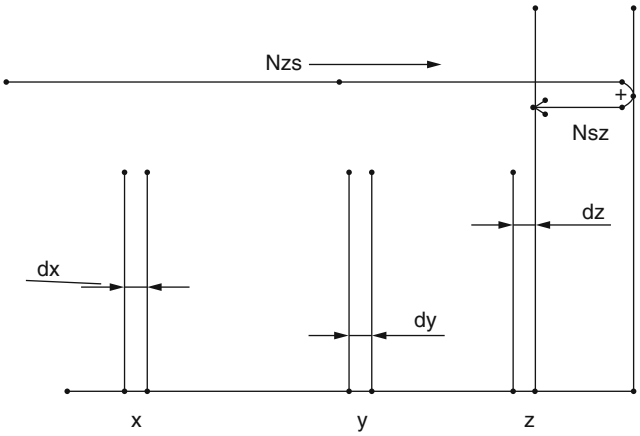


Fig. 7.68 Electron movement at the metal surface vicinity (z=0 is the surface)

takes place between electrons and phonons through the material, can be assessed by the total collision probability of electrons. Probability of electron traveling a distance z, where $z \parallel 2\lambda$, (λ being the mean free path), without making a collision is $\exp\left(-\frac{z}{\lambda}\right)$, the probability of an electron having just collided in dz is dz/λ.

In general, thermal conductivity is usually defined with respect to steady heat transfer through homogeneous medium in a random process. Thus, the thermal conductivity in one direction can be defined as:

$$k = \frac{N\overline{V}_z k_B \lambda}{6} \tag{7.362}$$

Note: that N is the number of free electrons per unit volume. It has been shown that for the temperatures higher than the Debye temperature, thermal conductivity can be assumed constant. The number of the electrons which have just collided in section I is: $N_{\zeta z} A_z \overline{V}_z \frac{d\zeta}{\lambda}$ Therefore, the number of electrons, which have just collided in $d\zeta$ during dt is:

$$N_{\zeta z} A_z \overline{V}_z \int_{-\infty}^{+\infty} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) \frac{d\zeta}{\lambda} \frac{dz}{\lambda} \tag{7.363}$$

and which then travel to section II before colliding in dz, where $\int_{-\infty}^{+\infty} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) \frac{d\zeta}{\lambda} \frac{dz}{\lambda}$ is the total electrons-lattice site atoms collision probability as described in [57].

The negative sign of the integral is due to a mirror image introduced at a surface. This mirror image represents reflected electrons from the free surface. The net

transfer of energy during the electron–phonon collision through the entire body can be written as:

$$\Delta E_{z,t} = \int_{-\infty}^{+\infty} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) \frac{dz}{\lambda} \frac{d\zeta}{\lambda} (f) \cdot [E_{\zeta,t} - E_{z,t}] \tag{7.364}$$

where $E_{\zeta,t}$ and $E_{z,t}$ are energy of electrons and photons at a considered region respectively, in the z-axis and parameter f is a fraction of electron energy. It is suggested that the rate of transfer of energy between the electrons and the molecules will be determined only by the difference in temperature of the electrons and lattice vibrations. If the temperature of the lattice site atoms in $u(v,T) = N(v)\langle E\rangle = \frac{8\pi v^2}{c^3}$ $\langle E\rangle$ is $\Theta(\zeta,t)$ then the energy transfer to the lattice site atoms in dz from collisions with electrons in which the electrons give up a fraction "f'" of their excess energy is

$$N_{\zeta z} A_z \overline{V}_z \frac{d\zeta}{\lambda} \frac{dz}{\lambda} \exp\left(-\frac{|z-\varsigma|}{\lambda}\right) f \cdot k_B [\Theta(\zeta,t) - \Theta(z,t)] \tag{7.365}$$

Summing the contributions from all such sections to obtain the energy in section *II* gives

$$\Delta E_{z,t} = A_z dz dt \int_{-\infty}^{+\infty} \frac{N_{\zeta z} \overline{V}_z f \cdot k_B}{\lambda} \frac{}{\lambda} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) [\Theta(\zeta,t) - \Theta(z,t)] d\zeta \tag{7.366}$$

During electron–phonon collision, some fraction f of the electron excess energy is transferred to the phonon. For any inelastic collision, the conservation of energy in any section may be written as:

$$\begin{aligned}\text{Electron energy entering the section} = {} & \text{Electron energy entering the section} \\ & + \text{Energy transfer to photons in the section}\end{aligned} \tag{7.367}$$

This gives:

$$f = \frac{\text{Electron energy in} - \text{Electron energy out}}{\text{Excess electron energy in}} \tag{7.368a}$$

$$f = \frac{(E_{eI})_{in} - (E_{eI})_{out}}{(E_{eI})_{in} - E_{photon}} \tag{7.368b}$$

providing that $0 \leq f \leq I$ from the energy conservation, where $(E_{eI})_{excess} = (E_{eI})_{in} - E_{photon}$ and E_{photon} = mean energy of phonon. The effective f value over a region sufficiently large to allow many collisions approaches unity and this corresponds to attainment of thermal equilibrium. In the case of a single collision, f depends only upon the masses of the colliding particles, according to

$$f = \frac{2M \cdot m}{(M+m)^2} \tag{7.369}$$

where M and m are the masses of an atom and an electron, respectively. Substituting appropriate values shows, that the f value is of the order of 10^{-4}. The most energetic phonon is only 0.01 eV, assuming $c_s = 10^5\,\mathrm{cm/s}$ (velocity of sound in the solid), but electrons near the Fermi level have energies of several eV; hence, when such electrons are scattered, only a small fraction of their energy can be given during an electron–phonon collision. In the present analysis, f is equal to 10^{-4} and is assumed as a constant over successive collisions. The change of irradiance of the laser beam passing through a homogeneous medium (metal), as a function of distance is given by $dI_z/d_z = -\mu I_z(z)$, where μ is the absorption coefficient. The negative sign indicates the reduction in beam irradiance due to absorption as μ is a positive quantity. The absorption of the incident laser beam takes place in z-axis. Integrating defined dI_z/d_z, the intensity of the incident beam at any plane z inside the substrate is given by $I_z = I_0\exp(-\mu z)$, where I_0 the peak intensity of incident irradiance. The limit for a small section Δz wide at z, the energy absorbed.

$$\frac{dI_z}{dz} = \frac{d}{dz}(\mu I_0 \exp(-\mu z)) \tag{7.370a}$$

or

$$I_z = -I_0 \frac{d}{dz}|f(z)| \tag{7.370b}$$

or

$$I_z = -I_0 f^{'}(z) \tag{7.370c}$$

where $f'(z)$ is the absorption function. Since the mirror image situation is considered, Fig. 7.68, for electron movement at the surface, the laser beam will be absorbed in a manner described by: $-f^{'}(z) = \frac{d}{dz}|\exp(-\mu|z|)|$. for all z. Using Eq. 7.368a and 7.368b for intensity of the laser beam, the rate of applied external energy at dz during the time interval dt can be given as

$$\Delta E_{z,t}|_{abz} = I_0 \mu e^{-\mu z} A dt dz \tag{7.371}$$

The total energy increases in the material at dz during dt is:

$$NA(E_{z,t+dt} - E_{z,t})dz = \Delta E_{z,t} + E_{z,t}|_{abs} \tag{7.372}$$

The total amount of energy, which is absorbed in an element $\partial\zeta$, area A in time dt is: $-I_0 A dt d\zeta f^{'}(\zeta)$, since all the beam energy is absorbed in the z-direction. One must

allow for the possibility that electron densities may vary throughout the material and, in particular, the number traveling from $d\zeta$ to dz may not be the same as that from dz to $d\zeta$. Therefore, the proportion of energy which is absorbed by the electrons which travel from $d\zeta$ to dz in dt is: $-I_0 A dt d\zeta f'(\zeta) \frac{N_{\zeta z}}{N_{\zeta z}+N_{z\zeta}}$

The average energy absorbed by one electron in $d\zeta$ in a time dt is: $-I_0 \frac{f'(\xi)d\xi}{(N_{\zeta z}+N_{z\zeta})\overline{V}_z}$,

and the total amount absorbed by this electron from dz to $d\zeta$ is: $\int_z^\zeta I_0 \frac{f'(\xi)d\zeta}{(N_{\zeta z}+N_{z\zeta})\overline{V}_z}$

The final temperature of the electrons in dz after the collision process can be readily found from the conservation of energy, i.e.,

$$\begin{aligned}\text{Total electron energy after collision} = {} & \text{Total electron energy in during } dt \\ & - \text{change site energy}\end{aligned}$$

The assumption that all directions of travel are equally probable gives: $N_{\zeta z} = N_{z\zeta} = \frac{N}{6}$

where N is the number of free electrons per unit volume, therefore, we can write

$$\begin{aligned}\frac{\partial}{\partial t}(\rho c \Phi(z,t)) = {} & \int_{-\infty}^{\infty} \frac{fk}{\lambda^3} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) \Theta(\zeta,t) d\zeta \\ & - \int_{-\infty}^{\infty} \frac{fk}{\lambda^3} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) \Theta(z,t) d\zeta \\ & + \int_{-\infty}^{\infty} \frac{I_0 f}{\lambda^2} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) \int_z^\zeta f'(z) d\xi d\zeta \end{aligned} \tag{7.373}$$

and

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{fk}{\lambda^3} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) & [\Theta(\zeta,t)d\zeta - f\Phi(z,t)d\zeta] \\ = {} & \int_{-\infty}^{\infty} \frac{k}{\lambda^3} \exp\left(-\frac{|z-\zeta|}{\lambda}\right) (1-f)\Theta(\zeta,t) d\zeta \\ & + \int_{-\infty}^{\infty} (1-f) \exp\left(-\frac{|z-\zeta|}{\lambda}\right) \int_z^\zeta f'(z) d\xi d\zeta \end{aligned} \tag{7.374}$$

Equations 7.373 and 7.374 are of interest to laser machining. The method of solution to be used in the following analysis is the transformation of the simultaneous differential integral Eqs. 7.373 and 7.374 using the Fourier integral transformation, with respect to A. The resulting ordinary differential equations may then be handled much more conveniently. The Fourier transformation of a function $f(z)$ is defined [57] by

$$F[f(z)] = \int_{-\infty}^{\infty} \exp(-i\omega z) f(z) dz = F(\omega) \tag{375}$$

and the Fourier inversion is given as

$$f(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)[\exp(-\omega z) d(\omega)] = F(\omega) \tag{7.376}$$

The Fourier transformation of convolution integral $\int_{-\infty}^{\infty} f(\xi) g(z-\zeta) d\zeta$ is the product of the transforms $F(\omega)\overline{g}(\omega)$ and the transform of function $\exp\left(\frac{|z|}{\lambda}\right)$ is $\frac{2\lambda}{1+\omega^2\lambda^2}$. Applying Fourier transformation to Eqs. 7.371 and 7.372 yields

$$\left(f + \omega^2\lambda^2\right) \frac{\partial}{\partial t}(\rho c \Phi) = -\omega^2 k f \overline{\Theta} + I_0 \delta f \frac{2\delta}{\delta^2 + \omega^2} \tag{7.377}$$

The multiplication in the transform domain by $(i\omega)^2$ corresponds to a second order differential in the real plane. Hence, the inversion of Eq. 7.377 gives

$$\left(f - \frac{\lambda^2\mu^2}{\partial z^2}\right) \rho c \frac{\partial \Phi}{\partial t} = kf \frac{\partial^2 \Phi}{\partial z^2} + I_0 \mu f \exp(-\mu|z|) \tag{7.378}$$

If the term $\left(\frac{\lambda^2\mu^2}{f\partial z^2}\right)\rho c \frac{\partial \Phi}{\partial t}$ is neglected for all f values, Eq. 7.376 becomes

$$\rho c \frac{\partial \Phi}{\partial t} = k \frac{\partial^2 \Phi}{\partial z^2} + I_0 \mu \exp(-\mu|z|) \tag{7.379}$$

which is the same as a Fourier heat conduction equation, shows as Eq. 7.182. It is apparent that the electron kinetic theory equations for the heat conduction process are much more general than the Fourier equation. The new model of the conduction process is valid in regions close to the surface where an absorption process takes place and therefore, the temperature profiles, which are obtained in these regions, can be expected to be valid.

One further advantage of this new approach is that the problem is completely specified; together with spatial boundary conditions, by the final equations, and that these equations can be solved using the method of Fourier transformation.

The Eq. 7.377 in the transform plane for the Gaussian intensity pulse is:

$$\Phi(z,t)=\frac{AI_0\mu}{\rho c}\left[\frac{1}{\beta\left(1-\mu^2\lambda^2\right)+\kappa\mu^2}\right]\left\{\begin{array}{l} i\mu\sqrt{\frac{\kappa-\beta\lambda}{\beta}}\exp\left[-\frac{\beta\kappa t}{\kappa-\beta\lambda^2}\right]\exp\left[i|z|\sqrt{\frac{\beta}{\kappa-\beta\lambda^2}}\right] \\ \qquad \operatorname{erfc}\left[\frac{|z|}{2\sqrt{\kappa t}}+i\sqrt{\frac{\beta\kappa t}{\kappa-\beta\lambda^2}}\right] \\ -\exp\left[i|z|\sqrt{\frac{\beta}{\kappa-\beta\lambda^2}}\right]\operatorname{erfc}\left[\frac{|z|}{2\sqrt{\kappa t}}-i\sqrt{\frac{\beta\kappa t}{\kappa-\beta\lambda^2}}\right] \\ +\exp(\kappa\mu^2 t-\mu|z|)\operatorname{erfc}\left[\mu\sqrt{\kappa t}-\frac{|z|}{2\sqrt{\kappa t}}\right] \\ -\exp(\kappa\mu^2 t-\mu|z|)\operatorname{erfc}\left[\mu\sqrt{\kappa t}-\frac{|z|}{2\sqrt{\kappa t}}\right] \\ -2\exp\left[-\frac{\beta\kappa t}{\kappa-\beta\lambda^2}+\mu|z|\right] \end{array}\right\} \tag{7.380}$$

and gives the closed form result of the electron kinetic theory approach, which graphically is shown in Fig. 7.69.

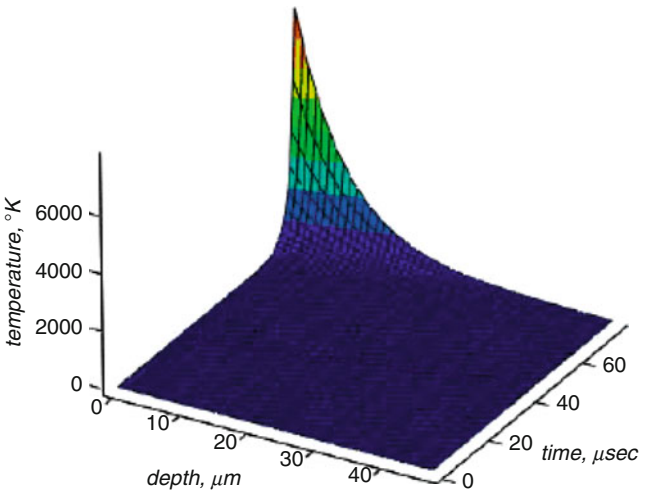


Fig. 7.69 Temperature, closed form solution of the electron kinetic theory approach, with respect to depth of the sample and time

7.10 Comparison of Fourier and Kinetic Theory

Comparison of analytical results of the Fourier relation given by Eq. 7.227, and kinetic theory equation, Eq. 7.321, shows that the two are identical when the following two conditions are meet: $\mu^2\lambda^2 \parallel 1$ and $\kappa \gg \beta\lambda^2$. For most materials, μ and λ are of the same order of magnitude. Since μ is of the order of 10^{-5} and λ is of the order of 10^{-8}, then $\beta \parallel 10^{11}$. This scenario corresponds to laser micromachining for laser pulses with picosecond rise times. Generally, $\beta \parallel \frac{\kappa}{\lambda^2}$ and Eq. 7.378 reduces to exactly the analytical solution obtained from the Fourier theory for the exponential pulse, provided that. $I \parallel \mu^2\lambda^2$. When "$\beta$" approaches zero, the pulse solution reduces to that for a constant intensity analytical solution of Fourier heating Eq. 7.227. The comparison of the temperature gradient $\partial\Phi/\partial z$ predicted from the Fourier and electron kinetic theories with the distance in the z-axis for two pulse lengths is shown in Fig. 7.70. The temperature gradients predicted from both theories are similar for the long pulse length $t_{\text{pulse}} = 6 \times 10^{-10}$. In general, the temperature gradient decreases sharply in the surface vicinity to reach its minimum.

As the distance from the point of minimum increases further inside the substrate, $\partial\Phi/\partial z$ increases gradually. The sharp decrease of $\partial\Phi/\partial z$ in the surface vicinity is due to the rapid increase of the temperature in this region. In this case, the energy absorbed by the electrons in the surface vicinity is converted into the internal energy gain of the substrate through collision process. This gives rise to a sharp increase of the lattice site temperature. The energy balance attains among the absorbed energy, internal energy gain, and the conduction process at the point of minimum $\partial\Phi/\partial z$. As the distance increases beyond the point of minimum, the

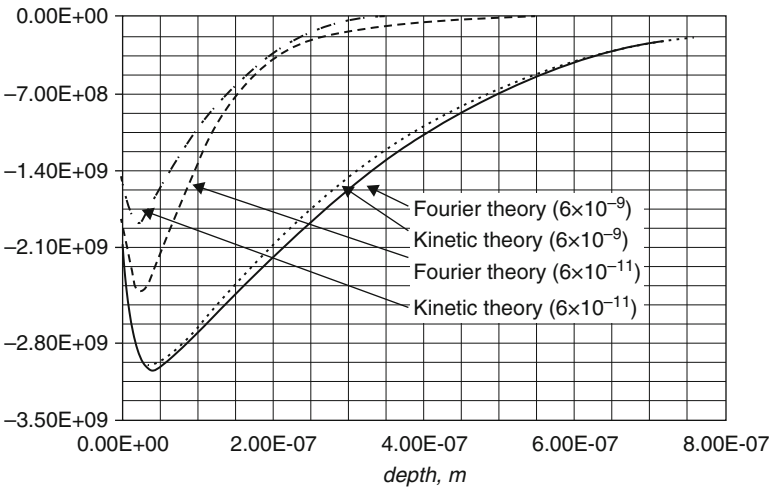


Fig. 7.70 dT/dz predicted from the Fourier and Kinetic Theory, along the -axis for two pulse lengths: 6×10^{-9}s and 6×10^{-11}s [57]

gradual increase in $\partial\Phi/\partial z$ reveals that the conduction effect due to phonon relaxation dominates. However, as the pulse length reduces, less $t_{\text{pulse}} < 6 \times 10^{-10}$ s, the temperature gradients predicted from both theories differ considerably. The temperature gradient predicted from the Fourier theory reduces significantly as compared to its counterpart predicted from electron kinetic theory. The difference in $\partial\Phi/\partial z$ predicted from both theories is due to the temperature response of the material for a short laser heating pulse as indicated earlier.

In summary, one can deduct a following conclusion: Temperature profiles predicted from Fourier heating model with kinetic theory approach for 1-dimensional model were compared in this study. In general, temperature profiles for 1-dimensional Fourier heating case and kinetic theory approach are very similar. Although the Fourier heating model fails to predict correct temperatures for the short pulse heating case, the equilibrium time exists for a given material, and the balance occurs between the internal energy gain due to laser irradiation and the conduction losses. The analytical solution of the electron kinetic theory approach obtained for the exponentially decaying pulse reduces to a step input intensity solution when β approaches zero. Moreover, the closed form solution of the electron kinetic theory for the step point intensity approaches the Fourier solution when $\mu^2 \frac{\lambda^2}{f} \leq 1$. The electron kinetic theory results deviate considerably from the Fourier theory results when $\beta \parallel 10^{11}$.

The difference in temperature profiles occurs because of the fact that the electrons in the surface vicinity absorb the incident laser energy and the excited electrons do not make sufficient collisions with the lattice site atoms to transfer their excess energy in the surface region. Thus, lattice site temperature in this region becomes lower than the electron temperature as evident from Fig. 7.70, in which the electron temperature distribution inside the substrate is shown. Therefore, the Fourier theory fails to predict the temperature rise in the surface vicinity accurately for heating time of $t_{\text{pulse}} < 6 \times 10^{-10}$s.

7.11 Finite Difference Methods

In this section, the solution of governing equation is accomplished using an explicit finite difference approximation and the corresponding boundary conditions. The representative results for computational investigations are presented and discussed. To solve general governing equation, Eq. 7.124, Finite Difference Method (FDM) was used, and the approximations to model the governing equation were based on forward-difference in time and central-difference in space. For more detailed information please refer to Nowakowski [40].

7.12 Effects of Pulsed Wave Laser Radiation

In this section and related sub-sections, we are going to deal with power levels of pulsed laser, Material vaporization effects and finally, the effects from absorption of radiation in the plume which we did touch before.

7.12.1 Power Levels of Pulsed Wave Laser

Highly intense pulses of short duration can be produced in a variety of ways. Typically it is done by creating a large population inversion by the injection of electrical energy from the discharge of large, highly charged capacitors. In these systems all the energy is produced in a burst, the duration of which can be made quite short. We shall not discuss the various techniques by which these pulses are created, but in Table 7.5 we simply note some commonly obtained values (Ready) [2].

From the above table it is apparent that with beam areas of the order of 1 cm^2 extremely high-power densities can be obtained, and, although the pulse lengths are short, the total energy in each pulse is considerable. The available power densities range as high as 10^{12} W/cm^2.

Practically speaking, one is usually interested only in power densities below the breakdown threshold of air because at higher power densities the energy never reaches the target. These breakdown levels are functions of wavelength, spot size, and pulse length, and depend as well on the contaminants in the air. Typical values are 10^9 W/cm^2 in "clean" air at Standard Temperature and Pressure (STP) for CO_2 laser pulses with duration of about 10^{-6} s and longer. At shorter pulse lengths the threshold is somewhat higher, becoming 10^{10} W/cm^2 at 10^{-8} s and 10^{11} W/cm^2 at 10^{-10} s. In the infrared region, the breakdown threshold scales with the square of the frequency.

7.12.2 Material Vaporization Effects

We shall first discuss the effect of high-power-density laser pulses on materials from the point of view of target vaporization, and shall assume that the vaporizing surface is not shielded from the radiation by the vapor. In this case we can show

Table 7.5 Some commonly obtained values for different pulsed lasers

Laser type	Pulse length	Power	Energy per pulse (J)
Ruby (normal mode)	0.1 to ms	1–100 kW	1–50
Ruby (Q-switched)	10^{-8} s	1–10 GW	1–10
Ruby (mode-locked)	10^{-11} s	0.1–1 TW	0.1–1
CO_2 TEA	10×10^{-6} s	100 MW	100
CO_2 e-beam	2×10^{-5} s	50 MW	1000
CO_2 shock tube gdl	3×10^{-4} s	0.3 MW	100

that, in addition to thermal input to the target, there is a strong pressure built up on the target surface due to recoil from the blow-off of the vapor. The integral of this pressure over the time of the laser pulse imparts a net impulse to the target. There arises then the possibility of inducing stresses large enough to create gross mechanical changes, such as spall and deformation, by pulsed laser irradiation.

To calculate the pressure applied to a surface by a laser pulse, we start with a consideration of the vaporization process. We use a one-dimensional calculation because in most cases of interest the beam radius R is larger than the thermal diffusion length during the pulse time t_p, or

$$R \gg 2\sqrt{\kappa t_p} \tag{7.381}$$

We shall avoid consideration of thin targets, so that ℓ is also large compared to the diffusion length. In this case we can calculate the time t_b required for the front surface to reach the vaporization temperature T_v from the semi-infinite slab result of Eq. 7.161, which is

$$T_v = \frac{2F_0}{K}\sqrt{\frac{\kappa t_b}{\pi}} \tag{7.382a}$$

or

$$t_b = \frac{\pi K^2 T_v^2}{4F_0^2 \kappa} \tag{7.382b}$$

or, since $\kappa = \frac{K}{\rho c}$, then Eq. 7.382b yields:

$$t_b = \frac{\pi}{4}\frac{K\rho c T_v^2}{F_0^2} \tag{7.382c}$$

In applying Eq. 7.161 in this way, we ignore the molten layer and assume that the values of K, ρ, and c appropriate to the solid can be used. This is not as gross an approximation as it may seem, because at these power densities the molten layer is very thin.

Once the material on the surface reaches the boiling point, the surface begins to erode at a rate U_s given by energy consideration, as we saw in Sects. 7.8 and 7.9 and their subsequent sub-sections as:

$$U_s = \frac{F_0}{\rho_s\left[c_{solid}T_m + L_m + c_{liquid}(T_v - T_m) + L_v\right]} \tag{7.383}$$

To simplify the calculation we take $c_{solid} = c_{liquid} = c$ and ignore L_m by comparison to L_v, then Eq. 7.383 reduces to:

$$U_s = \frac{F_0}{\rho_s(L_v + cT_v)} \tag{7.384}$$

Here we have used ρ_s, for the density of the solid. So after the time t_b given by Eq. 7.382c the surface begins to evaporate, and it recedes at the rate U_s. By conservation of momentum it must be true that

$$U_v\rho_v = U_s\rho_s \tag{7.385}$$

where ρ_v and U_v designate the density and velocity, respectively, of the evaporation products. Thus, we have:

$$U_v\rho_v = \frac{F_0}{L_v + cT_v} \tag{7.386}$$

by combining Eqs. 7.384 and 7.386.

To see how the pressure exerted on the surface is related to density and velocity, note that the pressure on the surface is just the pressure of the evaporation products. To calculate this pressure, consider particles which move a distance Δz in time Δt under the pressure P and thereby acquire a velocity V. The pressure (force per unit area) must equal the rate of change of momentum (per unit area) so that

$$P = \frac{(\rho\Delta z)V}{\Delta t} \tag{7.387}$$

That is, $\rho\Delta z$ is the mass per unit area which is brought to velocity V in time Δt by the force per unit area P. $V = \Delta z/\Delta t$, and so $P = \rho V^2$. Thus, in our specific case of density ρ_v and velocity U_v there is an associated pressure, given by:

$$P = \rho_v U_v^2 \tag{7.388}$$

We could compute the pressure from this expression if we knew ρ_v and U_v. However, we only know the product $\rho_v U_v$, from Eq. 7.386. We need another relationship, which we simply take from the ideal gas law,

$$P = \rho_v \frac{R}{A} T_v \tag{7.389}$$

where R is the gas constant and A the molecular weight. Denote R/A by C' and use Eq. 7.389, then

$$\rho_v U_v^2 = \rho_v C' T_v \tag{7.390}$$

or

$$U_v = \sqrt{C' T_v} \tag{7.391}$$

Upon combining Eq. 7.385 through 7.389, we get the desired relationship,

$$P = \frac{F_0 \sqrt{C' T_v}}{L_v + cT_v} \tag{7.392}$$

Since the specific heat of metal is typically $3R/A$ we can approximate C' by $(1/3)c$ to yield

$$P = \frac{1}{\sqrt{3}} \frac{F_0 \sqrt{cT_v}}{L_v + cT_v} \tag{7.393}$$

Finally, we compute the specific impulse delivered during the pulse, which is the force per unit area multiplied by the time over which it acts, and we get

$$I_{\mathrm{m}} = P\left(t_{\mathrm{p}} - t_{\mathrm{b}}\right) \tag{7.394}$$

or

$$I_{\mathrm{m}} = \frac{1}{\sqrt{3}} \frac{F_0 \sqrt{cT_v}}{L_v + cT_v} \left(t_{\mathrm{p}} - \frac{\pi}{4} \frac{K\rho c T_v^2}{F_0^2} \right) \tag{7.395}$$

A word about units is in order. It has become conventional to quote impulse in units of dyne-s, and specific impulse in dyne-s/cm^2. If we use J/cm^2 for energy density, J/g °C for specific heat, and J/g for heat of vaporization, we have

$$\begin{aligned} I_{\mathrm{m}} &= \frac{\mathrm{J/cm^2}}{\sqrt{\mathrm{J/g}}} \\ &= \mathrm{cm^{-2}} \sqrt{\mathrm{J\text{-}g}} \\ &= \mathrm{cm^2} \sqrt{10^7 \mathrm{erg} \times \mathrm{g}} \end{aligned} \tag{7.396}$$

or

$$I_{\mathrm{m}} = \sqrt{10^7} (\mathrm{dyne\text{-}s})/\mathrm{cm^2} \tag{7.397}$$

This unit, dyne-s/cm^2, is called a tap. Thus, in taps

$$I_{\mathrm{m}} = 1.83 \times 10^3 E_0 \frac{\sqrt{cT_0}}{[L_v + cT_v]} \left(1 - \frac{\pi}{4} \frac{K\rho c T_v^2}{E_0^2} t_{\mathrm{p}} \right) \tag{7.398}$$

Note that Eq. 7.398 predicts a threshold value of E_0 for impulse production at a given pulse length t_p. This is due to the criterion we introduced for vaporization; vaporization must commence before the end of the pulse or there will be no impulse. The threshold is given by

$$\frac{\pi}{4}\frac{K\rho c T_v^2}{E_0^2} t_p = 1 \tag{7.399}$$

or

$$E_0|_{th} = \frac{\sqrt{\pi}}{2} T_v \sqrt{K\rho c t_p} \tag{7.400}$$

This vaporization model also predicates that, at very large E_0, the impulse per unit area is directly proportional to the energy density with a constant *coupling coefficient*, given by:

$$\left(\frac{I_m}{E_0}\right)_{max} = 1.83 \times 10^3 \frac{\sqrt{cT_v}}{(L_v + cT_v)} \tag{7.401}$$

This is the limit at which vaporization begins essentially instantaneously with respect to the pulse length and vapor produces are produced for the entire pulse.

Some numerical values art, illustrated below and in Fig. 7.71. E_0 is in J/cm^2 and t_p in μs, so that I_m is in taps:

For Titanium

$$I_m = 8.04 E_0 \left[1 - (6.23)\frac{t_p}{E_0^2}\right] \tag{7.402}$$

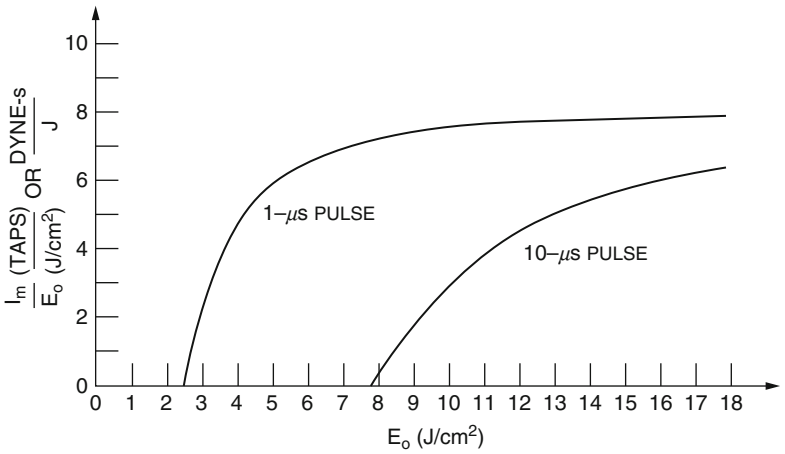


Fig. 7.71 Coupling coefficient vs. energy density for titanium

For Aluminum

$$I_{\mathrm{m}} = 6.49E_0\left[1 - (33.9)\frac{t_{\mathrm{p}}}{E_0^2}\right] \tag{7.403}$$

The above, model illustrates the principles involved in generating impulse by laser vaporization. In fact, in predicting threshold values it gives results which are within a factor of two of experimental measurements. It has been refined (Anisimov) [59] by a calculation which accounts for the fact that T_v is probably not a handbook value that comes from measurements at atmospheric pressure, but rather a different value appropriate to the dynamic and high-pressure situation created by the laser-induced vaporization. In this refinement, T_v is determined from the kinetic model of vaporization, which predicts that $U_{\mathrm{s}} = c_{\mathrm{a}}\exp\left[L_v/\left(R'T_v\right]\right.$ where the speed of sound in the solid is c_{a} and R' the gas constant per gram. When this is done, the thresholds agree very well with theory. However, as E_0 (and hence F_0, since t_{p} is constant) is increased, experiments show that delivered impulse does not increase indefinitely but to fall off. This is due to the onset of absorption of laser energy by the vapor products and/or the heated air near the target. We turn now to a consideration of this problem.

7.12.3 Effects from Absorption of Radiation in the Plume

The plume of vaporized material blown off the target becomes, at some power density hot enough that it or the air begins to absorb the laser radiation. The onset of this process is not thoroughly understood, and the ignition of these so-called absorption waves is then subject of a great deal of research. Proper treatment of the problem depends on, among other things, computing the onset of ionization and the rate of absorption of light by the electrons; and also accounting for both cascade processes and relaxation processes in a full dynamic sense. We shall not treat this problem here. Rather we shall look at some crude models which show, in a semi-quantitative way, various features of the absorption process.

First note that the decoupling of the absorption from the material surface due to shielding by the plume depends on the wavelength of the radiation. Recall from Eq. 7.100a that at the plasma frequency of the ν_{p} reflectivity of a "free-electron" metal drop sharply from a value near unity to essentially zero. If we assume that the coupling to the plume is due to the light interacting with electrons. Eq. 7.100a is valid. Using the mass of the free electron gives:

$$\nu_{\mathrm{p}} = 8.97 \times 10^3 N^{1/2} \tag{7.404}$$

For the plasma frequency ν_{p} in hertz, when N is in electrons per cubic centimeter. This can be rewritten in terms of the corresponding wavelength λ_{p} to yield:

$$N = \left(1.12 \times 10^{13}\right)/\lambda_p^2 \tag{7.405}$$

where λ_p is in centimeters. At a given wavelength the plume is transparent until the electron density reaches the value given by Eq. 7.405, where there will be a transition to a condition in which the plume absorbs and reflects the radiation and thereby shields the material. For 10.6 μm CO_2 radiation, shielding begins at 10^{19} electrons per cm^3, for 1.06 μm (Nd Laser) at 10^{21} electron per cm^3, and for 0.6943 μm (Ruby Laser) at 2.3×10^{21} electrons per cm^3.

When the electron density reaches a high enough value, the beam decouples from the surface and presumably the pressure due to material blow-off will drop. To get some idea of the order of magnitude of the energy density for a given pulse length where this process begins, let us simply assume that cutoff begins when the front surface reaches the temperature at which the material is fully ionized. 's should predict an upper limit, for full ionization is obviously not required. For example, solids have $N \approx 10^{23}$ cm^{-3}, whereas we only require, at 10.6 μm, $N > 10^{19}$ cm^{-3}. For simplicity assume that melting and vaporization processes can be ignored, and again use for the front surface temperature raises the simple expression

$$T = \frac{2F_0}{K}\sqrt{\frac{\kappa t}{\pi}} \tag{7.406}$$

A typical ionization temperature for a metal would be about 75,000 °C. Using simply the values of K and k for the solid, we get for titanium,

$$F_0\sqrt{t} \approx 5.7 \times 10^4 \mathrm{Ws}^{1/2}/\mathrm{cm} \tag{7.407}$$

Using $E_0 = F_0 t$, this can be rewritten as:

$$E_0 \approx 57\sqrt{t_{\mu s}} \tag{7.408}$$

where $t_{\mu s}$ is understood to be time in microseconds.

Figure 7.72 shows some data taken b. 1Dr. Rudder of the Air Force Weapons Laboratory, at two pulse lengths, 1.2 and 11 μs, with 1.06-pm radiation and titanium targets (Canavan et al. [61]). The lines marked "Anisimov predictions" are calculated from the vaporization model of the previous section with the refined method for determining T_v. (This was first done in the Soviet Union by (Anisimov [55]). The experimental data agree very well at values of E near threshold. Note that Eq. 7.406 for estimating the onset of shielding is roughly consistent with these data, although the experimental onset of shielding is, as one might expect, fairly gradual. The line on the graph marked LSD. Predictions refers to a theoretical estimate based on the idea that the laser light, when it couples into the blow-off, can create an explosion like shock wave in the air which travels up the beam, absorbing the radiation energy in the process. This Laser Supported Detonation, or LSD), wave is one form of laser-supported-absorption wave. We discuss these waves next.

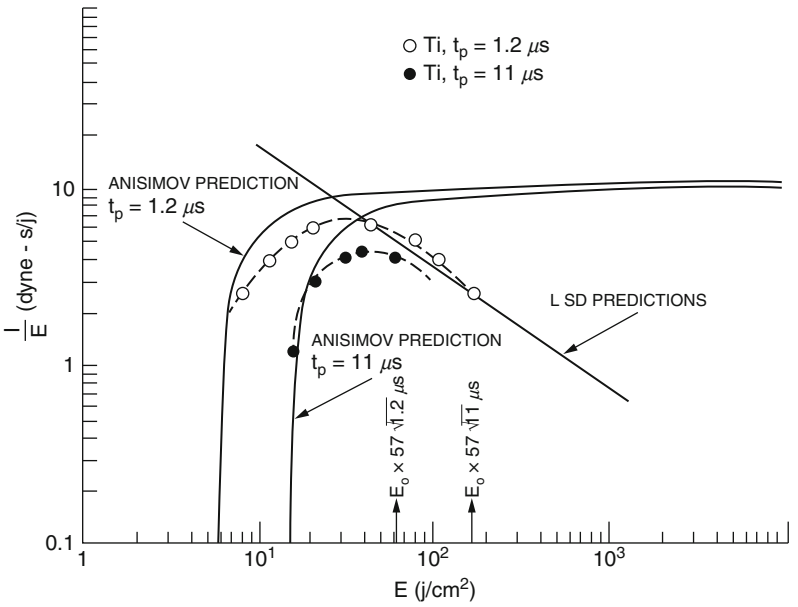


Fig. 7.72 Specific impulse delivered to solid targets by 1.06 μm laser radiation [61]

Once the coupling of the radiation with the ejected vapor (and perhaps the air) reaches a sufficient level, the absorption region begins to behave in a fashion characterized by hydrodynamic dissuasion of the energy coupled into it. For now, let us ignore the ignition problem. The absorption region typically propagates up the laser beam in a way that is determined by the medium in which it propagates (usually air) and also by the balance between the power being fed in by the laser and the relaxation processes which dissipate the power. Three types of laser-supported-absorption waves are usually identified. Typical power level at which they appear and their typical velocities of propagation are indicated below for 10.6 μm radiation and targets in air at standard temperature and pressure [61].

Type of wave	Power level of laser flux (W/cm^2)	Velocity of propagation (cm/s)
Laser supported detonation wave LSD	10^7	10^5
Laser supported combustion wave LSCW	10^4	10^3
Plasmatron	10^4	0

The LSD wave propagates as a shock wave, i.e., at supersonic velocity, whereas the LSC wave moves more slowly and relaxes by thermal conduction. The plasmatron is at rest, with the energy input being balanced by reradiation and convective losses into the atmosphere. Although we discuss these effects here in the section on pulsed lasers, they are just a.4 valid for continuous radiation. Since pulsed lasers are the, most convenient devices for reaching these power levels,

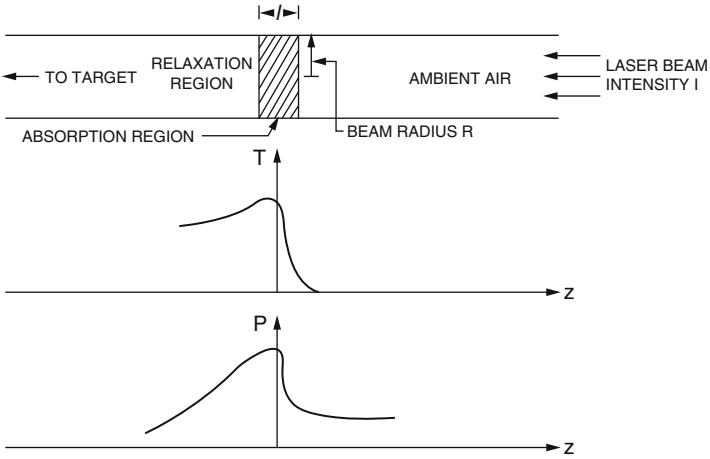


Fig. 7.73 Temperature and density profiles typical of a laser-supported detonation wave [60]

especially for LSD waves, absorption waves are usually considered under pulsed effects

Hydrodynamic theory can he applied Lo model these waves. The problem was first solved in the Soviet Union by (Raizer) [62]. Detonation waves can be discussed most. Readily because the hydrodynamic equations reduce- to fairly simple expressions, so we shall consider them in some detail. A few remarks about combustion waves will come later.

We can derive conditions for the steady-state behavior of a detonation wave by considering conservation of mass, momentum, and energy at the detonation front. For this purpose we do not concern ourselves with how the process starts but presume that a detonation wave has been formed and is propagating at some steady rate as sketched in Fig. 7.73. The absorption region is propagating to the right at a steady velocity u. We assume that it is very thin and can be treated as a detonation front. Thus, u is the detonation velocity. The temperature, density, etc. of the air go through very rapid changes in the very short distance ℓ. Note that this wave propagates, in this treatment in air, and thus, our results will be independent of target material

In this discussion "behind the front" refers to the high-temperature-and-pressure region immediately to the left of the absorption region in Fig. 7.73. "Ahead of the front" is to the right in the sketch and refers to ambient air conditions. Note we have given the beam a finite radius R and thus will have to consider lateral expansion. First let us do the one-dimensional problem and assume that the detonation front propagates simply as a plane wave.

Behind the front let ρ, P, and e be the density, pressure, and internal energy *per unit mass*, respectively, and ρ_0, P_0, and e_0 be the same variables ahead of the front. Define the velocities with respect to a coordinate system moving with the front at

the detonation rate u. Then the ambient gas moves into the front with the speed of u, and we define v as the speed with which the high-pressure gases leave the front. We can now write down the conservation equations for mass, momentum, and energy across the detonation front. These equations are based on *flow*, that is, they are in the terms of "per unit mass area, per unit time." The equation for mass is:

$$\text{Mass} \quad \rho_0 u = \rho v \tag{7.409}$$

The conservation-of-momentum condition results from the impulse to the change in momentum. Now impulse is force multiplied by time, but in the "per unit area, per unit time" sense this becomes simply pressure. Since mass, in this flow concept, is ρv, momentum is $(\rho_v)v = \rho v^2$. Hence, we have

$$P - P_0 = -[(\rho v)v - (\rho_0 u)u] \tag{7.410}$$

and, if we ignore P_0, which is much smaller than P, we have for momentum:

$$\text{Momentum} \quad P + \rho v^2 = \rho_0 u^2 \tag{7.411}$$

The conservation-of-energy condition follows from similar considerations. The difference in energy flow on each side of the front must be balanced by the work done on the gas $(P_0 u - Pv)$ and the energy absorbed from the laser beam, which, using our earlier notation, is F. $F = AI$, where A is the absorptance of the gas in the absorption region. Thus, we have:

$$\rho u\left(e + \frac{1}{2}v^2\right) - \rho_0 u\left(e_0 + \frac{1}{2}u^2\right) = P_0 u - Pv + F \tag{7.412}$$

If we use $P_0 \approx 0$ and $e_0 \approx 0$ and substitute from Eq. 7.409, we get for energy

$$e + \frac{1}{2}v^2 - \frac{1}{2}u^2 = -\frac{p}{\rho} + \frac{F}{\rho_0 u} \tag{7.413}$$

or

$$\text{Energy} \quad e + \frac{P}{\rho} + \frac{1}{2}v^2 = \frac{1}{2}u^2 + \frac{F}{\rho_0 u} \tag{7.414}$$

Our goal is to use these conservation laws to predict the pressure P behind the front and ultimately the pressure transmitted to the target. For now assume that F is known, and, of course, the ambient air density ρ_0 is known. Thus, we have three equations and five unknown, P, ρ, v, e and u. To proceed we need to invoke some equation of state for the gases, and we shall simply assume that the ideal gas law holds. Thus, we have:

$$P = \rho R' T \tag{7.415}$$

where R' is the gas constant per unit mass, or $R^{'} = R/\mathfrak{M}$, where $\mathfrak{M}$ is the molecular weight. Since the wave is presumed to be in air in this treatment, $\mathfrak{M}$ would be the average molecular weight of air. Taking $\mathfrak{M}$ for air to be 29.4/mol gives $R^{'} = 2.84 \times 10^6 \text{erg}/(g\,^\circ\text{C})$ and is consistent with the ideal gas law and with $\rho_0 = 1.26 \times 10^{-3}$ g/cm^3 for air at 0 °C and 1 atmosphere = 10^6 dyne/cm^2). Now Eq. 7.415 essentially introduces another unknown, the temperature T, so we need to add the expression for the energy of an ideal gas, which is:

$$e = \frac{R^{'}T}{\gamma - 1} = \frac{P}{(\gamma - 1)\rho} \tag{7.416}$$

Where γ is the ratio of the specific heat, $\gamma = c_p/c_v$. For our purpose it is sufficient to take $\gamma = 1.4$.

Now we have four equations that are Eqs. 7.407, 7.409, and 7.414 as well as Eq. 7.425 in five unknowns that are P, ρ, v, e and u. To get the final condition we use the criterion for detonation, which is the velocity of the high-pressure gases behind the front, relative to the front, is equal to or greater than the local speed of sound. Intuitively this seems reasonable, for propagation of shock waves is, by definition, in excess of the speed of sound. The criterion can be properly derived from a consideration of the thermodynamics of the situation, but we shall not do so here, however you can refer to (Ya. B. Zel'dovich and Yu P. Raizer) [63]. Since we shall be interested in the minimum value of F (or I) required to sustain in detonation wave, we take v equal to the speed of sound. For an ideal gas the sound speed is $(\gamma P/\rho)^{1/2}$, so we have our last condition as:

$$v^2 = \frac{\gamma P}{\rho} \tag{7.417}$$

Before discussing the algebra, let us collect the Eqs. 7.409, 7.411, 7.414, 7.425, and 7.417 respectively:

$$\begin{aligned}
&\text{Mass} \quad \rho_0 u = \rho v \\
&\text{Momentum} \quad P + \rho v^2 = \rho_0 u^2 \\
&\text{Energy} \quad e + \frac{P}{\rho} + \frac{1}{2}v^2 = \frac{1}{2}u^2 + \frac{F}{\rho_0 u} \\
&\text{Energy of Ideal Gas} \quad e = \frac{P}{(\gamma - 1)\rho} \\
&\text{Velocity Behind Shock} \quad v^2 = \frac{\gamma P}{\rho}
\end{aligned}$$

Combine Eqs. 7.409 and 7.411 to yield expressions for u^2 and v^2 as:

$$u^2 = \frac{P\rho}{\rho_0(p - \rho_0)} \tag{7.418}$$

$$v^2 = \frac{P\rho}{\rho(p-\rho_0)} \tag{7.419}$$

Now use Eqs. 7.415 and 7.417 to eliminate v^2 and ρ so that we can get:

$$\frac{\rho}{\rho_0} = \frac{1+\gamma}{\gamma} \tag{7.420}$$

which is one of the equations we need, namely ρ in terms of the known quantities ρ_0 and γ. Now we can use Eq. 7.411 to get P in terms of u by eliminating v^2 with Eq. 7.417 to get:

$$P = \frac{\rho_0 u^2}{1+\gamma} \tag{7.421}$$

We need one more relation to complete the solution, namely u in terms of F. This will, by Eq. 7.421, give us P in terms of F. To get this we use Eq. 7.414 and replace e via Eq. 7.416. Thus, Eq. 7.414 becomes:

$$\frac{P}{(1+\gamma)\rho} + \frac{P}{\rho} + \frac{1}{2}\frac{\gamma P}{\rho} = \frac{1}{2}u^2 + \frac{F}{\rho_0 u} \tag{7.422}$$

If we use Eq. 7.421 to eliminate P, we get

$$\frac{u^2}{1+\gamma}\left[\frac{1}{(\gamma-1)}\frac{\rho_0}{\rho} + \frac{\rho_0}{\rho} + \frac{1}{2}\frac{\gamma\rho_0}{\rho}\right] = \frac{1}{2}u^2 + \frac{F}{\rho_0 u} \tag{7.423}$$

or

$$\left(\frac{\rho_0}{\rho}\right)\left(\frac{u^2}{1+\gamma}\right)\left[\frac{\frac{1}{2}\gamma(1+\gamma)}{\gamma-1}\right] = \frac{1}{2}u^2 + \frac{F}{\rho_0 u} \tag{7.424}$$

or

$$\frac{1}{2}\frac{\rho_0}{\rho}u^2\left(\frac{\gamma}{\gamma-1}\right) = \frac{1}{2}u^2 + \frac{F}{\rho_0 u} \tag{7.425}$$

If we use Eq. 7.420 for ρ/ρ_0, we get

$$\frac{1}{2}u^2\left(\frac{\gamma}{\gamma-1}\right) = \frac{1}{2}u^2 + \frac{F}{\rho_0 u} \tag{7.426}$$

Finally we arrive at

$$u^3\left(\frac{1}{\gamma^2-1}\right) = \frac{2F}{\rho_0} \tag{7.427}$$

or

$$u = \left\{ \frac{2(\gamma^2 - 1)F}{\rho_0} \right\}^{1/3} \tag{7.428}$$

The equations which represent the solution for the detonation wave, then, are

$$\frac{\rho}{\rho_0} = \frac{1+\gamma}{\gamma} \tag{7.429}$$

$$P = \frac{\rho_0 u^2}{1+\gamma} \tag{7.430}$$

$$u = \left[\frac{2(\gamma^2 - 1)F}{\rho_0} \right]^{1/3} \tag{7.431}$$

These three equations, together with the ideal gas law, represent the formal solution to the propagation of the laser-supported-detonation wave. Given the temperature behind the front, and since Eq. 7.429 defines ρ, we could calculate P and hence u and finally the F required to support it. However, this does not really solve the problem. What we wish to discover is: given the laser intensity I, will an LSD wave be supported? To answer this question, we need to consider the distance it takes for the laser radiation to be absorbed. We also need a more realistic situation than the simple plane wave.

First we note that the beam has a finite radius R and that lateral expansion can take place. The order of magnitude of the radial expansion velocity will be the speed of sound c_a. To maintain the detonation, we must replace the energy lost to expansion by energy put into the absorption region. To simplify, let us assume all of the laser beam energy is absorbed in the distance ℓ. (Actually the beam intensity only falls by $1/e$ in the distance ℓ). We define Δt as the time for the shock front to move a distance ℓ, or $\Delta t = \ell/u$. In this time the radial expansion is the amount $c_a \Delta t$. Now $\pi R^2 I \Delta t$ is the energy deposited by the beam in the cylindrical volume shown in Fig. 7.74 ($c_a \Delta t << R$). But the energy in this volume after expansion is

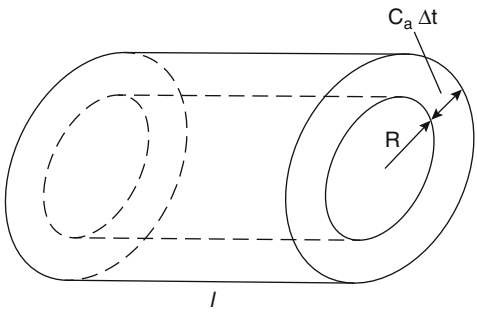


Fig. 7.74 Cylindrical volume which absorbs beam energy via expansion

approximately equal to its volume multiplied by its internal energy p, in unit volume. Thus

$$\pi R^2 I \Delta t \approx \rho_0 e \left[\pi R^2 \ell + 2\pi R c_a \Delta t \ell\right] \tag{7.432}$$

Since $\Delta t = \ell/u$.

$$I\pi R^2 \approx \rho_0 e u \left[\pi R^2 + 2\pi R c_a \frac{\ell}{u}\right] \tag{7.433}$$

Note that $\gamma \approx 1.4$, Eq. 7.429 predicts that $\rho \approx 2\rho_0$ and hence, using $\rho v = \rho c_a \equiv \rho_0 u$, $u \approx 2c_a$. So we have, after some algebra

$$\rho_0 e u \approx \frac{I}{1 + \ell/R} \tag{7.434}$$

But this equation simply represents the rate at which we must put energy into the absorption volume in order to maintain the conditions we assume to carry out our detonation wave calculation, namely a plane wave propagating by absorption of laser energy in a distance ℓ. Thus, the energy flow per unit area from the laser beam is:

$$F = \frac{I}{1 + (\ell/R)} \tag{7.435}$$

Finally, we can complete the problem if we know the absorption length ℓ. To compute ℓ we need to invoke some model of the ionized air. For this purpose it is sufficient to assume that free electrons absorb the light and that the electrons come from singly ionized atoms. We shall not derive the expressions which we need but simply quote them. There are two relationships. The first of these is that is known as Saha Equation [63] Eq. 7.425 below, which relates the fraction of atoms ionized α, to the absolute temperature T and the ionization potential J of a single atom:

$$\frac{\alpha^2}{1-\alpha} = 2\frac{g_1}{g_0}\frac{m}{\rho}\left(\frac{2\pi m_0 kT}{h^2}\right)^{3/2} e^{-J/kt} \tag{7.436}$$

In this equation m is the mass of the atom, m_0 is that of the electron, k is Boltzmann's constant, and h is Planck's constant. The statistical weights of ground state of the atom and its first ionized state are g_0 and g_1, respectively. Typically $g_1 = g_0 = 1$. In terms of known constants, then, the Saha equation gives us the degree of ionization as a function of temperature.

Knowing the degree of ionization, we can get the absorption length. Again we simply quote the relationship [60], which assumes that the light is absorbed by inverse Bremsstrahlung. The expression is

$$\frac{1}{\ell} = \frac{4}{3}\left(\frac{2\pi}{3m_0kT}\right)^{1/2} \frac{e^6h^2}{m_0c(h\nu)^3}\left(\frac{\rho^2\alpha^2}{m^2}\right)\left(1 - e^{-h\nu/hT}\right) \tag{7.437}$$

which, at the temperature of interest and for 10.6 μm radiation, becomes (for $h\nu/h$ $T \parallel 1$):

$$\frac{1}{\ell} = \frac{4}{3}\left(\frac{2\pi}{3m_0}\right)^{1/2} \frac{e^6h^2}{m^2m_0c(h\nu)^2} \cdot \frac{\rho^2\alpha^2}{(kT)^{3/2}} \tag{7.438}$$

In Eq. 7.438, ν is of course, the frequency of the laser radiation, c is the speed of light, and e is the electronic charge. By combining Eqs. 7.437 and 7.438 we can calculate ℓ in terms of temperature T, density ρ, and known parameters. A typical value of J for air is of the order of 13 eV. For O_2, J is 12.1 eV, for N_2 it is 15.6 eV, and thus, we have ℓ as a function of ρ and T as:

$$\ell = f(\rho, T) \tag{7.439}$$

or, since $\rho = \rho_0(1+\gamma)/\gamma \approx 2\rho_0$, we can get a relationship between ℓ and T and hence between I and T, via Eqs. 7.431 and 7.435. Typically results are shown in Fig. 7.75.

Here we have assumed a beam radius of 10 cm. The important point is that there is a *minimum* in the I vs. T relation. We identify this as the minimum flux I_m required to maintain LSD wave. Associated with it is the temperature T_m of the high-pressure region at the detonation front.

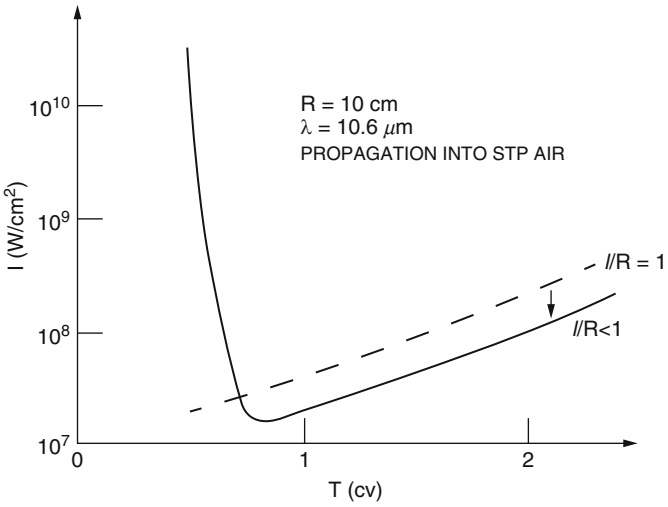


Fig. 7.75 Relationship between I and T for a laser-supported detonation wave [60]. This plot is based on a more realistic expression for the equation of state of the gases than the ideal gas law, but the ideal gas law gives a similar result

We can then use our detonation-wave relationships, Eqs. 7.429 and 7.431 through Eq. 7.420, to get the pressure behind the front, or equivalently, get pressure from T_m via $\rho \approx 2\rho_0$ and ideal gas law.

We shall turn to a calculation of the pressure on the target in a moment. First note that the radial expansion concept imposes a natural criterion for the difference between a combustion wave and a detonation wave. The time for radial expansion is R/c_a, whereas the time for passage of the absorption region is V/u. If the detonation condition is to be maintained, radial expansion times must be larger than propagation times in order for the high-pressure region to move as a shock front and not dissipate itself radially. Hence, $R/c_a > \ell/u$. We have already noted that $u \approx 2c_a$ crudely, $u \approx c_a$, so that $R > \ell$, or $\ell/R < 1$, is the condition for detonation waves. If ℓ becomes larger than R, the absorption region is large, the relaxation in the radial direction is important, and the process called a combustion wave takes place. This can be treated in a similar fashion to the detonation wave, but the hydrodynamic equations do not take the simple form of Eqs. 7.409, 7.411, and 7.413. We shall not treat combustion waves in this book. The solution in Fig. 7.75 is for a detonation wave and hence is valid for $\ell/R < 1$. The limit $\ell/R = 1$ is shown in the figure by a dashed line.

Finally, compute the impulse delivered to target by a laser beam of intensity I just sufficient to maintain a detonation wave [61]. The beam has a pulse duration t_p. We wish to calculate the effect on the target due to the "explosion products" behind the absorption region. These, of course, expand in all directions and create a pressure on the target. To demonstrate the effect we shall use a very simple model, namely a model of cylindrical expansion. We consider that the absorption region has propagated a distance Z by the end of the laser pulse, and at that time we have created a cylinder of high pressure gas which has a radius equal to the beam radius R, a length Z, and a pressure P_d given by Eqs. 7.430 and 7.431 above, with $P = P_d$. This cylinder is then allowed to expand radially at a speed estimated to be the speed of sound c_a. Then we get the impulse delivered to the target by integrating the force on the target due to the pressure in the expanding cylinder during the time the cylinder expands from R to the target radius R_T. For R_T very large, the integration is stopped when the cylinder pressure drops to atmospheric pressure. The model is sketched in Fig. 7.76 at the time $t = t_p$.

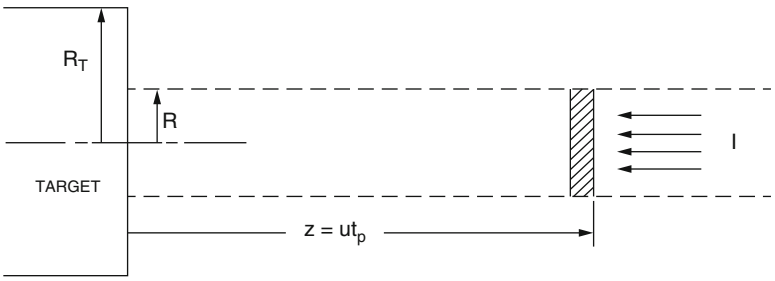


Fig. 7.76 Radial expansion model for impulse delivered to a target from a laser-supported detonation (LSD) wave

The model might be expected to be valid if $Z \ll R$ and if $t_p \ll$ the time required for radial expansion, either to R_T or to atmospheric pressure, to take place. We also are assuming that impulse due to target vaporization is negligible, i.e., that the detonation wave is formed very early in the laser pulse.

We take the radial expansion to be at constant temperature. Then P times the volume V, of the cylinder is a constant. Since our model presumes only cylindrical expansion, we have the condition that.

$$Pr^2 = \text{Constant} \tag{7.440}$$

where r is the radius of the cylinder and $R \leq r \leq R_T$. We shall need this relationship in the derivation of the impulse. Let the impulse be I_m, and let F be the force on the target due to the pressure. This gives the following:

$$I_m^{'} = \int_{t(r=R)}^{t(r=R_T)} F dt \tag{7.441}$$

where the upper limit is understood to be valid only where P is greater than atmospheric pressure at $r = R_T$. Since the radial expansion rate is c_a. From ideal gas relation for the speed of sound, we have:

$$c_a = \sqrt{\frac{\gamma P}{\rho}} \tag{7.442}$$

we have

$$dt = \sqrt{\frac{\rho}{\gamma P} dr} \tag{7.443}$$

Thus

$$I_m^{'} = \int_R^{R_T} F \sqrt{\frac{\rho}{\gamma P}} dr \tag{7.444}$$

Now at any time

$$F = (\pi r^2) P = \pi r^2 P \tag{7.445}$$

but, since $r^2 P$ is constant, then we can evaluate F from the initial pressure P_d at $r = R$, or

$$F = \pi R^2 P_d \tag{7.446}$$

Now the impulse becomes, by taking F out of integral sign of Eq. 7.444, then we have

$$I_{\mathrm{m}}^{'} = \pi R^2 P_{\mathrm{d}} \int_R^{R_{\mathrm{T}}} \sqrt{\frac{\rho}{\gamma P}} dr \tag{7.447}$$

again invoking $r^2 P = \text{constant}$, gives

$$r^2 P = P_{\mathrm{d}} R^2 \tag{7.448}$$

or

$$P = P_{\mathrm{d}} \frac{R^2}{r^2} \tag{7.449}$$

So the impulse is then:

$$I_{\mathrm{m}}^{'} = \pi R \sqrt{P_{\mathrm{d}}} \int_R^{R_{\mathrm{T}}} \sqrt{\frac{\rho}{\gamma}} r dr \tag{7.450}$$

Recall from Eq. 7.429 that $\rho = \rho_0(1+\gamma)/\gamma$. This gives

$$I_{\mathrm{m}}^{'} = \pi R \sqrt{\frac{P_{\mathrm{d}} \rho_0 (1+\gamma)}{\gamma^2}} \left(\frac{R_{\mathrm{T}}^2}{2} - \frac{R^2}{2} \right) \tag{7.451}$$

Since $R < R_{\mathrm{T}}$, we shall ignore the $R^2/2$ term. Thus, specific impulse I_{m}, which is $I_{\mathrm{m}}^{'}$ divided by the area of the beam πR^2 is then

$$I_{\mathrm{m}} = \frac{I_{\mathrm{m}}^{'}}{\pi R^2} = \frac{R_{\mathrm{T}}^2}{2R} \sqrt{P_{\mathrm{d}} \rho_0} \frac{\sqrt{1+\gamma}}{\gamma} \tag{7.452}$$

Recalling the expression in Eqs. 7.430 and 7.431 for P_{d}, and ignoring ℓ/R with respect to unity, we can write:

$$I_{\mathrm{m}} = \frac{R_{\mathrm{T}}^2}{2R} \sqrt{\left(\frac{1+\gamma}{\gamma^2} \right) \rho_0 \left(\frac{\rho_0}{1+\gamma} \right) \left[\frac{2(\gamma^2 - 1)}{\rho_0} I \right]^{2/3}} \tag{7.453}$$

This simplifies to:

$$I_{\mathrm{m}} = \frac{R_{\mathrm{T}}^2}{2R} \left\{ \frac{1}{\gamma} \left[2(\gamma^2 - 1) \right]^{1/3} \right\} \rho_0^{2/3} I^{1/3} \tag{7.454}$$

The expression in braces is nearly equal to unity for typical vale of γ (say, $\gamma = 1.4$), so, Eq. 7.454 reduces to:

$$I_{\mathrm{m}} = \frac{R_{\mathrm{T}}^2}{2R}\rho_0^{2/3} I^{1/3} \tag{7.455}$$

Now the energy per unit area in the beam is $E_0 = It_{\mathrm{p}}$, and we write the coupling coefficient I_{m}/E_0 as:

$$\frac{I_{\mathrm{m}}}{E_0} = \frac{R_{\mathrm{T}}^2}{2R}\rho_0^{2/3} / \left(t_{\mathrm{p}}^{1/3} E_0^{2/3}\right) \tag{7.456}$$

This is the equation of the straight line marked Laser Supported Detonation (LSD) Predictions shown in Fig. 7.72, where the calculations were done for the parameters appropriate to the 1.2 μs pulse length.

Several important consequences of the LSD wave are seen in Eq. 7.445. One is that the coupling coefficient is reduced as E_0 becomes larger, which tells us that we cannot create an arbitrarily large impulse at a target by simply increasing the energy in the laser beam. In fact, when Eq. 7.456 is considered to be correct at high E_0 and the results of the vaporization model (see Eq. 7.398 and Fig. 7.72) are used at lower values of E_0, there is, for a given pulse length, an optimum value of E_0 for transferring the largest amount of impulse to a target. For the 1.2 μs pulse illustrated in Fig. 7.72 the optimum value of E_0 is about 22 J/cm^2, and this is in reasonable accord with the data. Of course the specific impulse I_{m} *per se* goes as $I_{\mathrm{m}}/E_0^{1/3}$. so larger E_0 will create larger I_{m}. However, this slow increase of 1 m with E_0 is a very inefficient way to impart stress to a material. A better scheme, perhaps, would be to use multiple pulses at the optimum E_0 value.

Another consequence of the LSD wave is a lack of dependence of impulse on the parameters of the target material. The same impulse is produced independently of the target. This is in accord with experiment. When I is well into the range where LSD's are formed, the measured impulse is the same for all target materials. Some data taken by Naval Research Laboratory (NRL) [64] are shown in Fig. 7.77. In this graph we see, in accord with the vaporization model, a strong dependence of I_{m}/E_0 on material type at lower power densities, whereas at the high-power densities typical of LSD formation the values of I_{m}/E_0 are the same for all materials. In this range, however, a target area dependence appears. The target area dependence shown is for aluminum. Here again the general behavior predicted by Eq. 7.445 can be seen.

We have not yet considered the radius at which the expanding cylinder reaches atmospheric pressure. Call this radius R_0. As explained above Pr^2 is constant, so

$$P_0 R_0^2 = P_{\mathrm{d}} R^2 \tag{7.457}$$

where P_0 is atmospheric pressure, 106 dyne/cm^2. Thus,

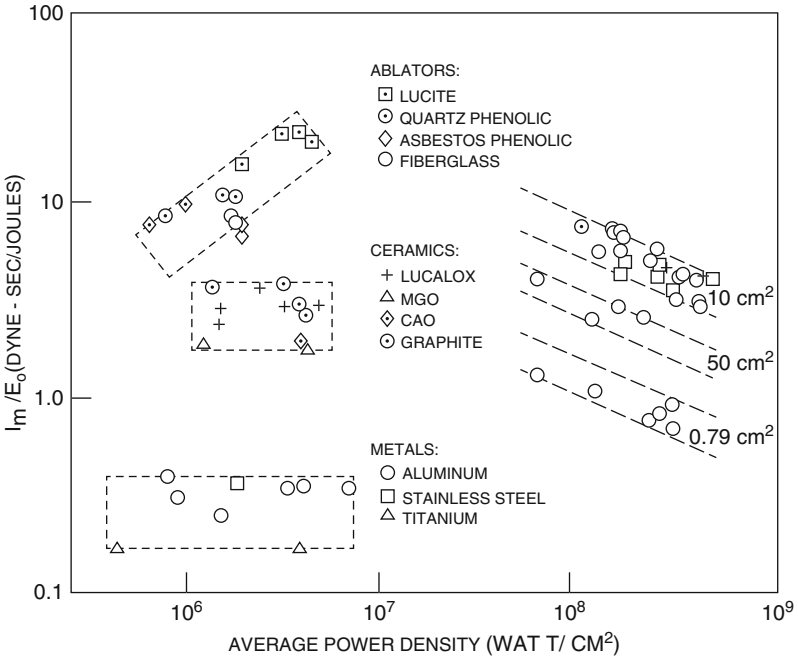


Fig. 7.77 Coupling coefficient as a function of power density [64]

$$R_0 = RP_0^{-1/2} P_d^{1/2} \tag{7.458}$$

From Eqs. 7.419 and 7.420, for P_d this becomes

$$R_0 = RP_0^{1/2} \sqrt{\frac{\rho_0}{1+\gamma} \left[\frac{2(\gamma^2-1)t}{\rho_0}\right]^{2/3}} \tag{7.459}$$

Again the factor involving γ are nearly unity, so

$$R_0 = RP_0^{1/2} \rho_0^{1/6} I^{1/3} \tag{7.460}$$

Upon substituting $I = E_0/t_p$, we get

$$R_0 = RP_0^{1/2} \rho_0^{1/6} E_0^{1/3} t_p^{-1/3} \tag{7.461}$$

So if R_0 is less than R_T, one replace R_T by R_0 in Eq. 7.456. For target sizes larger than R_0 will all receive the same impulse.

A numerical illustration is useful. Let $E_0 = 1000\,\mathrm{J/cm^2}$ and $t_p = 100\,\mu\mathrm{s}$. Suppose the beam radius is 1 cm. The power density I is about 10^7 W/cm^2, so we expect a

LSD wave. If we take $R_T = 5\,\text{cm}$ for the target radius, Eq. 7.445 yields (with $\rho_0 = 1.29 \times 10^{-3}\,\text{g/cm}^3$)

$$\frac{I_m}{E_0} = \frac{25}{2}\left(1.29 \times 10^{-3}\right) / \left[\left(10^{-4}\right)^{1/3}\left(10^{10}\right)^{2/3}\right]\text{dyne-s/erg}$$

$$\frac{I_m}{E_0} \approx 7 \times 10^{-7}\,\text{dyne-s/erg}$$

or

$$\frac{I_m}{E_0} \approx 7\,\text{dyne-s/J}$$

In this example expansion to atmospheric pressure would take place at a radius given by Eq. 7.461:

$$R_0 = \left(10^6\right)^{-1/2}\left(1.29 \times 10^{-3}\right)^{1/6}\left(10^{10}\right)^{1/3}\left(10^{-4}\right)^{-1/3}$$

Where we have used erg/cm^2 for E_0. Then

$$R_0 \approx 15\,\text{cm}$$

Thus, targets with radii of 15 cm and larger would exhibit a maximum coupling coefficient of $(15/5)^2$ time 7, or 65 dyne-s/J. If we compute I_m, we have 63,000 dyne-s/cm^2, that is 63,000 taps, as the maximum specific impulse from this laser pulse. Since this impulse is delivered in times of the order of magnitude of the laser pulse time, this corresponds to a pressure of roughly $\left(6.3 \times 10^4\right)/10^{-4} \approx 6 \times 10^8\,\text{dyne/cm}^2$, or about 600 atmospheres.

7.13 Effects of Continuous Wave Laser Radiation

The choice of laser system depends on the application at hand. First of all the structuring of metals, semiconductors, and dielectrics is very diverse. Since ultrashort pulse lasers operate at very high intensities, they can employ nonlinear absorption and are therefore an obvious choice for laser processing of transparent materials, but it all depends on the specifications of the application.

Laser can be categorized most easily according to their mediums, which are divided into three basic categories as defined by the state of the lasing material:

(a) Gas
(b) Liquid
(c) Solid

Furthermore, all laser types operate in one or two temporal modes:

1. Continuous wave (CW) mode
2. Pulse mode

In CW mode, the laser beam is emitted without interruption. In pulsed mode as we describe in previous sections, the laser beam is emitted periodically.

If structures on the tens of micrometers scale are the aim of the application, a nanosecond laser will often provide sufficient accuracy, and will typically have a greater throughput, thereby making it the preferred solution. If, however, features on the nanometer scale are required, a picosecond or femtosecond laser system would be a better choice.

For nanosecond laser systems the resolution of the generated features is limited by the heat diffusion length, $\ell \approx \sqrt{\kappa\tau}$, where κ is the thermal diffusivity of the material and τ is the laser pulse duration [65]. Metals typically have a high thermal diffusivity, and precise micro-structuring must therefore be undertaken with shorter pulses [66, 67]. It should also be noted that nanosecond laser processing is associated with a heat-affected zone, i.e., an area around the laser-generated hole where the temperature has been high enough for the material to undergo melting (but not vaporization). Though the material resolidifies, the strength of this area may have been greatly reduced.

The output of a laser can be continuous wave (CW) or pulsed. In pulsed operation, much higher peak powers can be achieved since the energy stored in the laser gain medium is released in a short burst. Laser ablation is the process of removing material from a surface by laser beam irradiation. Short pulse laser ablation is advantageous since the material can be heated up to the temperature of vaporization in a very short time. This means that the energy does not have time to spread into the deeper parts of the material and thereby the energy is localized where it is needed.

For a sufficiently low power flux the thin surface layer will be heated to the fluid state but will stay beyond the evaporation temperature, while the solid–fluid interface will slowly progress into the bulk material by heat conduction. In iron the typical progress rate is about 10^{-2} cm m/s, and the power flux for this situation may be around 10^6 W/cm^2.

At a somewhat higher power flux, between 10^6 and 10^8 W/cm^2, the thin absorbing surface layer is heated up to its evaporation temperature before the solid–fluid interface has progressed appreciably into the material by heat conduction. Thus, with continuing laser power, a less than μ-thick layer of material will continually evaporate, with the material–air interface progressing into the material. Typically a gas jet develops. The gas jet ejects also part of the molten material, and thus, the progress rate of the hole is faster than with evaporation of all the material.

At a still higher flux of 10^9–10^{10} W/cm^2, after initial evaporation of the surface layer the gas jet is thermally ionized and absorbs most of the incident radiation, which is blocked away from the material. The surface layer explodes with an ultrasonic jet; its temperature may rise beyond 10^5 K, its pressure beyond 10^3 at material properties or select from a set of predefined materials.

As we say at the beginning of this chapter, light contains energy. When that light is absorbed by a material, its energy goes toward heating the material. However, the faster the material gains heat, the faster it sheds heat. If the rate of heat loss gets high enough to equal the rate of heat gain, the death ray can no longer raise the temperature of the surface. Otherwise, the surface continues to gain heat and its temperature rises. The temperature to which a beam of light can bring a material depends on the intensity of the light and the rates and mechanisms by which the material responds to the heat.

Generally, there are three mechanisms by which heat moves around: heat conduction, heat convection, and radiation. Heat conduction is what you get when heat moves through a material, transferred from atom to atom by the atoms bumping into each other and exchanging energy. Convection is the movement of heat by moving the material it is stored in. Radiation transfers heat by the emission or absorption of electromagnetic waves. When a surface is exposed to a death ray, it gets heated by radiation.

For continuous wave (CW) laser beams of relatively low intensity, the only effect is heating from absorption of electromagnetic radiation. The temperature of the absorbing material is raised until a balance exists between the absorbed power and thermal losses from conduction, convection, and radiation.

It has been known since prehistory that sunlight can make an absorbing surface hot, and in Greek antiquity it was known that concentrated sunlight, having traversed spherical water-filled flasks or convex pieces of glass, could kindle fires[3]. Archimedes proposed to concentrate the sun's rays by means of reflecting the harbor of Syracuse. He hoped to ignite the ropes, sails, and spars of the vessels of enemy's fleet.

The power flux density of the sun at zenith on the earth's surface is about 0.13 W/cm^2 which leads to the solar furnaces with temperatures near 3000 °C. If the irradiated area is large enough, lateral heat conduction may be ignored. The radiative loss of a black-body surface is given by Stefan–Boltzmann's law. If the irradiated area is large enough, lateral heat conduction may be ignored. The power flux density, σT^4, corresponds to 1 kW/cm^2 at $T = 3644$ K. Convective cooling by air-flow over a 3000 K surface at Mach number unity is only a few hundred W/cm^2. At pressures prevailing at booster burn-out altitudes, 80–160 km, convective cooling is completely negligible. From the foregoing, it is clear that the temperature of most materials may be raised above the melting temperature T_m and the vaporization temperature T_v for CW laser flux densities in the range of 1–100 kW/cm^2. In the early days of laser history, in 1961, when the pulsed ruby laser was the most powerful available, it was established that a focused ruby laser pulse of about 1 J energy could punch a hole in a razor blade[4]. Two very simple cases serve to

[3] *Aristophanes, Comedy of Clouds,* 434 B.C., English translation

[4] The strength of those pulses was measured in unofficial unit of "Gillette."

establish the order of magnitude of fluxes and fluencies on target required for lethality [68]. Most analysis that we have showed so far are applicable for the CW laser type as well.

References

1. Schriempf JT (1974) Response of materials to laser radiation a short course. NRL Report 7728, July 10, 1974. Department of Navy, Office of Naval Research Arlington, VA 22217
2. Ready JF (1971) Effects of high-power laser radiation. Academic Press, New York
3. Carslaw HS, Jaeger JC (1959) Conduction of heat in solids, 2nd edn. Clarendon Press, Oxford
4. Nowakowski KA (2005) Laser beam interaction with materials for microscale applications. PhD Dissertation, Worcester Polytechnic Institute, 22 November 2005
5. Ross D (1969) Lasers: light amplifiers and oscillators. Academic Press, New York, p 72
6. Wester R (2011) In: Poprawe R (ed) Tailored Light 2: laser application technology. New York, Springer Publication
7. Born M, Wolf E (1999) Principles of optics. Cambridge University Press, Cambridge
8. Zettili N (2009) Quantum mechanics, concept and applications, 2nd edn. Wiley, New York
9. John F (1971) Ready, effects of high-power laser radiation. Academic Press, New York
10. Wieting TJ, Schriempf JT (1972) Free electron theory and laser interaction with metals. Report of NRL Progress, June 1972, pp 1–13
11. Bonch-Bruevich AM, Imas YaA (1967) Zh Tekh Fiz 37:1917 (English transl.: Sov PhysTech Phys 12:1407 (1968))
12. Zhang SY, Ren YH, Lupke G (2003) Appl Opt 42(4):715
13. Romero LA, Dickey FM (1996) J Opt Soc Am A 13(4):751
14. Momma C, Nolte S, Kamlage G, von Alvensleben F, Tunnermann A (1998) Appl Phys A Mater Sci Process 67(5):517
15. Sanner N, Huot N, Audouard E, Larat C, Huignard JP, Loiseaux B (2005) Opt Lett 30 (12):1479
16. Nemoto K, Nayuki T, Fujii T, Goto N, Kanai Y (1997) Appl Opt 36(30):7689
17. McLeod E, Hopkins AB, Arnold CB (2006) Opt Lett 31(21):3155
18. Heinemann S (1995) Opt Commun 119(5–6):613
19. Dickey FM, Holswade SC (eds) (2000) Laser beam shaping: theory and techniques. Marcel Dekker, New York
20. Bäuerle D (2000) Laser processing and chemistry. Springer, Berlin
21. Heller J, Bartha JW, Poon CC, Tam AC (1999) Appl Phys Lett 75(1):43
22. Toulemonde M, Unamuno S, Heddache R, Lampert MO, Hageali M, Siffert P (1985) Appl Phys A Mater Sci Process 36(1):31
23. Toulemonde M, Unamuno S, Heddache R, Lampert MO, Hageali M, Siffert P (1985) Appl Phys A Mater Sci Process 36(1):31
24. Arnold CB, Aziz MJ, Schwarz M, Herlach DM (1999) Phys Rev B 59(1):334
25. Weeber JC, Krenn JR, Dereux A, Lamprecht B, Lacroute Y, Goudonnet JP (2001) Phys Rev B 64(4):045411
26. Baeuerle D (2000) Laser processing and chemistry, 3rd edn. Springer, New York. ISBN 3-540-66891-8
27. Lide DR (2001) CRC handbook of chemistry and physics, 82nd edn. CRC, Boca Raton
28. Slusher RE, Eggleton BJ (2004) Nonlinear photonic crystals, 1st edn. Springer, Berlin
29. Ghofraniha N, Conti C, Ruocco G, Trillo S (2007) Phys Rev Lett 99(4):043903
30. Staudt W, Borneis S, Pippert KD (1998) Phys Status Solidi A Appl Res 166(2):743
31. Mori N, Ando T (1989) Phys Rev B 40(9):6175
32. Columbia (2005) Laser micromachining, 10 November 2005. www.mrl.columbia.edu

33. Lasag (1997) Operator's manual for KLS 126 laser source. LASAG Corporation, Switzerland
34. Semak V, Matsunawa A (1997) The role of recoil pressure in energy balance during laser materials processing. J Phys D Appl Phys 30:2541–2552
35. Bronski MT (2003) Development of a process characterization of Nd:YAG crystals. MS Thesis, Center for Holographic Studies and Laser micro-mechaTronics, Mechanical Engineering Department, Worcester Polytechnic Institute, Worcester, MA
36. Bonch-Bruevich AM, Imas YaA (1967) Zh Tekh Fiz 37:1917 (English transl.: Sov Phys Tech Phys 12:1407 (1968))
37. Nowak T (1990) Theoretical and experimental investigation of laser drilling in a partially transparent medium. MS Thesis, Center for Holographic Studies and Lasermicro-mechaTronics, Mechanical Engineering Department, Worcester Polytechnic Institute, Worcester, MA
38. Yilbas BS (1995) Study of liquid and vapor ejection processes during laser drilling of metals. J Laser Appl 7:147–152
39. Tokarev VN, Lunney JG, Marinea W, Sentis M (1995) Analytical thermal model of ultraviolet laser ablation with single-photon absorption in the plume. J Appl Phys 78(2):1241–1246
40. Yilbas BS, Yilbas Z, Akcakoyun N (1996) Investigation into absorption of the incident laser beam during Nd:YAG laser processing of metals. Opt Laser Technol 28(7):503–511
41. Yilbas BS, Yilbas Z, Sami M (1996) Thermal processes taking place in the bone during CO_2 laser irradiation. Opt Laser Technol 28(7):513–519
42. Tabor D (1991) Gases, liquids and solids and other states of matter, 3rd edn. Cambridge University Press, Cambridge, p 272
43. Bejan A (1993) Heat transfer. Wiley, New York, NY
44. Mazumder J, Steen WM (1980) Heat transfer model for cw laser material processing. J Appl Phys 51:941–947
45. Gordon R, Cobonpue J (1961) Heat transfer between a flat plate and jets of air impinging on it. Heat Transfer Pt. II:454–460, ASME, New York, NY
46. Niedrig R, Bostanjoglo O (1996) Imaging and modeling of pulse laser induced evaporation of metal films. J Appl Phys 81(1):480–485
47. Semak V, Damkroger B, Kemka S (1999) Temporal evaluation of the temperature filed in the beam interaction zone during laser material processing. J Phys D Appl Phys 32:1819–1825
48. Zohuri B (2015) Dimensional analysis and self-similarity methods for engineers and scientists. Springer Publishing Company, New York
49. Parker WJ, Jenkins RJ, Butler CP, Abbott GL (1961) J Appl Phys 32:1679
50. Heckman RC (1971) Thermal diffusivity finite pulse time corrections. Sandia Laboratories Research Report SC-RR-710280, May 1971
51. Schriempf JT (1972) Rev Sci Instrum 43:781
52. Sparks M (1975) Theory of laser heating of solids: metals. J Appl Phys 47(3):837–849
53. Yilbas BS, Sami M (1997) Liquid ejection and possible nucleate boiling mechanism in relation to the Laser drilling process. J Phys D Appl Phys 30:1996–2005
54. Sparks M (1975) Theory of laser heating of solids: metals. J Appl Phys 47(3):837–849
55. Anisimov SI, Khokhlov VA (1995) Instabilities in laser-matter interaction. CRC, Boca Raton, FL
56. Yilbas BS, Gbadebo SA, Sami M (2000) Laser heating: an electro-kinetic theory approach and induced thermal stresses. Opt Laser Eng 33:65–79
57. Yilbas BS, Arif AFM (2001) Material response to thermal loading due to short pulse laser heating. Int J Heat Mass Transfer 44:3787–3798
58. Nash GE, Thermal response calculation. Naval Research Laboratory, Washington, DC, unpublished data
59. Anisimov SI (1968) High Temp 6:110. Translated from Teplofizika Vysokikh Temperature 6:116 (1968). Original article submitted December 6, 1966
60. Nielsen PE, Canavan GH (1971) Laser absorption waves. Air Force Weapons Laboratory Laser Division Digest LRD-71-2, p 110, December 1971

61. Canavan GH, Nielsen PE, Harris RD (1972) Radiation of momentum transfer to solid targets by plasma ignition. Air Force Weapon Laboratory Laser Division Digest LRD-72-1, p 125, June 1972
62. Raizer YP (1965) Sov Phys JETP 21:1009
63. Zel'dovich YB, Raizer YP (1966) Physics of shock waves and high-temperature hydrodynamic phenomena. Academic Press, New York, p 346
64. Metz SM, Hettche LR, Stegman RL (1973) Intensity dependence of target response to high-intensity pulsed 10.6 μ laser radiation. In DOD Laser effects/hardening conference, Monterey, CA, 23–26 October 1973
65. Bäuerle D (2000) Laser processing and chemistry, 3rd edn. Springer, Berlin
66. Dausinger F, Lichtner F, Lubatschowski H (2004) Femtosecond technology for technical and medical applications. Springer, Berlin
67. Balling P (2006) In: Kane DM (ed) Laser cleaning II. World Scientific Publishing Company, Singapore, p 257
68. The American Physical Society (1987) Report to the American Physical Society of the Study Group on Science and Technology of Directed Energy Weapons. The American Physical Society

Chapter 8
Atmospheric Propagation of High-Energy Laser Beams

Starting from the invention of LASER in the early 1960s, laser radiation propagation in the atmosphere has been the subject of intensive research. The high spatial and time coherence of laser sources makes their application attractive for communication, location, geodesy, and high-energy transmission over long distances. Laser sources are widely used for exploring the atmosphere, in particular, its gas composition and pollution, velocities of air and sea flows, and features of the land and sea surface.

Incorporating HEL systems into military operations is not without challenge. The impact of the environment—in the atmosphere, over land, over water, and in space—on system performance can be significant. Understanding and predicting such impacts, as well as the effects and vulnerabilities of HEL systems, can be important considerations in designing systems, identifying promising areas of technology research, developing concepts of operations, and employing HEL systems on the battlefield.

8.1 Introduction

In this chapter we will discuss the response of materials to a high-power radiation laser with a one-dimensional mathematical modeling and presentation of solution to different cases of heat conduction partial differential equation along with a given boundary; a directed energy weapon (DEW) has been a recurring theme in science fiction literature and cinema ever since H.G. Wells published the "War of the Worlds" in 1898. The idea of a "death ray" which can instantly destroy or burn a target at a distance retains its allure to this very day. More than a century after Wells contrived his "heat ray," the technology is maturing to the point of becoming soon deployable.

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7_8

High-energy laser weapons have been progressively evolving since the 1960s, a path punctuated by a series of important scientific breakthroughs and engineering milestones.

The popular view of a HEL, seen as constructing a great big laser and pointing it at a target with the intention of vaporizing it, bears only vague similarity to a real HEL weapon. There are genuine technological and operational challenges involved in creating truly useful and effective weapons.

A laser weapon system uses high-energy lasers such as deuterium fluoride (DF) or hydrogen fluoride (HF) for which all necessary reactants and other materials are carried on an easily movable vehicle. Reactant gases are stored mixed with a diluent gas, such as helium, for ease of handling and to provide nearly ideal gas behavior. Cooling water for the laser is also employed in a high-pressure steam generator that uses diesel fuel and oxygen to produce heat. Apart from a fluorine generator, the system uses only four storage tanks for reactant gases.

But what are the advantages of using a laser as a weapon? Is it even possible? Could you use such a weapon to stun an opponent? These questions are being addressed by the Air Force Research Laboratory's Directed Energy Directorate. This program is developing high-energy lasers, microwave technologies, and other futuristic weapon systems, such as the airborne laser and the PHaSR.

Lasers and other directed energy weapons have many advantages over conventional projectile weapons like bullets and missiles:

- The weapons' light outputs can travel at the velocity of light.
- The weapons can be precisely targeted.
- Their energy output can be controlled—high power for lethal outcomes or cutting and low power for nonlethal outcomes.

The propagation of high-energy laser (HEL) beams in the atmosphere is rich in fundamental physics and of paramount importance to the Navy's directed energy research program. Laser beams with hundreds of kilowatts to megawatts of average power are affected by numerous interrelated linear and nonlinear phenomena such as molecular and aerosol absorption and scattering, atmospheric turbulence, and thermal blooming. Aerosol scattering and absorption are often the major limiting factors in HEL propagation. In particular, aerosol absorption has been shown to be a major factor leading to thermal blooming. Properties of aerosols are often found in outdated tables using methodologies that may not be consistent with HEL propagation. In particular, under nonlinear conditions attained by HELs, aerosols can change their scattering and absorption properties during the HEL engagement time. In a typical directed energy engagement scenario at multi-kilometer ranges, approximately half of the laser power can be lost to aerosol scattering. In addition, the maritime propagation environment is characterized by strong turbulence which causes beams to wander and spread. Standard adaptive optic (see Sect. 8.4) methods fail to compensate for the effects of deep turbulence, which can result in significant power loss at the target. Thus it is important to characterize and understand the interaction of HEL beams with aerosols and the effects of deep turbulence on beam propagation.

Early experiments showed that laser beams propagating in the atmosphere are distorted because of the turbulent inhomogeneity of the refractive index. Not far away from the source, the turbulence distorts the phase front of the light wave and disturbs the spatial coherence.

The effect of the atmospheric turbulence on laser propagation has been thoroughly studied [1–10]. The results suggest that adaptive optic systems [11–13] will eventually eliminate turbulence-caused noise.

At higher intensities or radiation powers of hundreds of kilowatts in the continuous-wave mode available today, laser beam propagation in the atmosphere is hindered by another source of distortion: inhomogeneities resulting from the radiation action on the medium [14–26]. One mechanism responsible for atmospheric air parameter variation under the action of radiation is heat. The air temperature and, consequently, its optical parameters vary as a result of radiation absorption by atmospheric gases and aerosols. Even in originally homogeneous media, the refractive index may vary either regularly (when the medium is heated by radiation with a regular intensity distribution) or randomly if the intensity is a random function of space and time.

Thus heat-induced changes in the optical parameters of the medium lead to changes in radiation propagation. This change is referred to as *thermal blooming*.

Thermal blooming is manifest as strong distortions of laser beams that were first detected in beam propagation in absorbing fluids. Calculated data suggested that similar phenomena could be observed also in the propagation of high-power radiation along lengthy atmospheric paths, though the radiation is but weakly absorbed by air. This conclusion promoted an interest in thermal blooming of laser beams in the atmosphere.

Hundreds of publications are now available which discuss different aspects of thermal blooming. Most publications deal with the thermal blooming in aerosols and clouds where thermal nonlinear effects become conspicuous at lower intensities and in greater variety than in the transparent atmosphere. The findings on the thermal blooming in aerosols are summarized in References [7, 19, 20].

Thermal blooming in a transparent atmosphere having low absorption and scattering factors is discussed in References [16–18]. These surveys are concerned mostly with the results of numerical and experimental simulation of the thermal blooming in homogeneous (in the absence of radiation) media.

A specific feature of the atmospheric propagation paths is regular or random variation of the medium parameters. In regular variations, the magnitude and direction of the wind velocity are critical for the blooming, while in random ones, turbulent pulsations of both wind velocity and medium temperature are vital. Some of the information here is compiled as the results of theoretical and experimental studies of thermal blooming carried out in the Institute of Atmospheric Physics of the USSR Academy of Sciences. The studies were aimed chiefly at forecasting the effects of thermal blooming in a turbulent atmosphere and testing the blooming calculation procedures in a turbulent medium by laboratory simulation. The combined effect of thermal nonlinearity and turbulence on the propagation of laser beams is discussed in Sect. 8.2 and its subsections.

The summary of such effort by USSR Academy of Sciences, which includes a theoretical study of the effect of turbulence on beam propagation, was carried out by solving nonlinear equations by various perturbation techniques. A better insight into the blooming process in a medium assumed homogeneous was required for the study. Specifically, the following points had to be clarified: applicability of an isobaric approximation to the calculation of the perturbations in a medium subjected to radiation and of specific features of thermal blooming in the modes when the medium compressibility cannot be neglected, applicability of various approximate techniques of thermal blooming analysis, the effect of the initial intensity distribution and regular velocity variation of the wind along the blooming path, and stability of blooming.

8.2 Laser Propagation in the Atmosphere

Atmospheric effects dictate the design and performance of almost all high-energy laser systems (SBLs being exceptions). These effects include those common to all electro-optical systems, namely, obstruction by opaque clouds, transmission losses from scattering and absorption, and optical turbulence degradation. They also include effects unique to HEL systems, such as thermal blooming arising from molecular and aerosol absorption.

High-energy laser beams propagating through the atmosphere can be severely defocused or deflected by thermal blooming [17]. The thermal blooming process is driven by a small fraction of the laser energy that is absorbed by the molecular and aerosol constituents of air [27, 28]. The absorbed energy locally heats the air and leads to a decrease in the air density which modifies the refractive index. The refractive index variation leads to a defocusing or deflection of the laser beam. In the presence of a transverse wind, the region of heated air is convected out of the beam path, and a steady-state situation is realized [17]. In general, however, the intensity of a beam undergoing thermal blooming is a function of both time and spatial position, particularly in a stagnation zone, where the effective wind velocity is zero.

Stagnation zones, i.e., regions in which the effective wind velocity is zero, can greatly enhance the thermal blooming of high-energy laser (HEL) beams in the atmosphere. An expression for the Strehl ratio of a focused HEL beam propagating through a stagnant absorbing region is derived. The propagation of a HEL beam in a maritime atmosphere is numerically modeled in a fully three-dimensional and time-dependent manner. The beam is focused onto a remote target, and a stagnation zone is created by slewing the laser in the direction of the wind. The laser power delivered to the target is calculated as a function of slew rate. For the parameters considered, it is found that a stagnation zone near the laser source has little effect on the propagation efficiency, while a stagnation zone near the target can significantly reduce the power on the target.

Stagnation zones are particularly detrimental to HEL propagation since, without an effective clearing mechanism for the heated air, the strength of the thermal lens grows in time. In this situation, the defocusing process is eventually limited by thermal conduction or buoyancy. However, by the time these processes become effective, the beam may have already been severely degraded.

Thermal blooming in the presence of a stagnation zone was experimentally observed in a laboratory experiment using a 10-W, CO_2 laser passed through an absorption cell containing CO_2 gas. The cell was pivoted to simulate a wind profile containing a stagnation zone. This experiment was also modeled using a code that solved time-dependent thermal blooming equations in the isobaric regime [29].

In the atmosphere, beam propagation and divergence are to a first approximation the same as in a vacuum, with the added feature that interaction of the beam with atmospheric constituents causes it to lose photons. The intensity of the beam then decreases with range for two reasons: divergence increases the beam size, and atmospheric interactions reduce the energy that it carries. Having already quantified the first of these effects, we will turn our attention to the second. Photons may be lost from the beam in several ways. They may be scattered or absorbed by atmospheric gases or particulate contaminants. They may be bent from the beam by the lensing effect of density fluctuations in the atmosphere. And at high intensities, they may cause the air through which the beam passes to break down into the absorbing plasma. These and related effects will be considered in this section [30].

Adaptive optics provides the means to maintain beam quality in the face of atmospheric turbulence. However, such methods are not applicable or completely effective in all situations. Ideal adaptive optics requires a beacon—a point source of light from the target—and this is not completely or even partially achievable in some systems or applications. Furthermore, for long slant or near-horizontal propagation paths, the integrated turbulence strength can be sufficiently strong that even the best adaptive optics cannot completely compensate for the turbulence. Thus the system performance is degraded by the atmospheric turbulence conditions and the limited capabilities of the adaptive optic system. Volume I of this book went through more details of adaptive optical system (AOS) and mirror conjugation as well as mirror juddering in Chap. 6.

System performance is significantly enhanced by a capability to model and predict laser system effectiveness under specific atmospheric conditions. Forecasting and decision aids for existing conventional electro-optical systems provide a model for HEL systems. A joint program, involving the Army, Navy, and Air Force, has developed and delivered decision aids for systems where atmospheric effects are a concern, including low-light-level TV systems, passive infrared seekers, and laser-guided munitions.

The program in atmospheric measurements and modeling for the airborne laser provides expanded understanding of atmospheric effects applicable to emerging HEL systems. High-altitude clouds (cirrus) and optical turbulence fundamentally limit ABL effectiveness and range.

Early in the program, it was recognized that the variability of turbulence was producing variability in ABL performance. As a result, a parallel Air Force Research Laboratory (AFRL) S&T (Scientific and Technical) program (with limited core S&T funding augmented by ABL funds) was initiated to examine atmospheric measurements and modeling. This effort has evolved into the Atmospheric Decision Aid (ADA) program.

Emerging HEL systems for air-to-ground applications suffer far more performance variability than does the ABL in its missile defense mission. This is a result of the atmospheric boundary layer and the degree to which weather and diurnal cycle affect performance. In addition to turbulence, the full range of cloud fields and aerosols will need to be modeled in air-to-ground applications.

The first-order effect of clouds on electro-optic (EO) and HEL systems has to do with the line of sight to the target. This effect is quantified through the cloud-free line-of-sight (CFLOS) statistic. Although CFLOS is a basic concept, it is somewhat unique to military problems. Additionally, clouds vary significantly with season and location, and therefore a firm understanding of the climatology of CFLOS over militarily significant areas is needed. As future HEL systems are specified and designed, realistic physical models of clouds and CFLOS are required—particularly in light of the increased importance of virtual engineering, simulation, and testing. And as these systems are deployed, the ability to forecast clouds in terms of CFLOS probabilities will become essential. These capabilities will require an improvement in the current ability to forecast clouds, including improvements in satellite cloud sensing, using numerical weather models.

Although employment modes for some HEL systems may eliminate or minimize atmospheric effects (such as the space-based laser or reduced distances to the target for the ABL), it is desirable that these systems be effective in much more broadly defined scenarios, especially when adjunct missions are considered. Atmospheric modeling and decision aids will significantly enhance HEL systems and expand their operational capabilities, much like the demonstrated contribution of atmospheric decision aids to the effectiveness of comparatively simpler systems such as infrared (IR) seekers. Like the ADA program for the ABL, expanded capabilities will need to be tailored to the operational scenarios and lethality mechanisms of new HEL systems. Advancements in atmospheric modeling and decision aids will require an expanded, long-term S&T program, as the need for atmospheric models is military specific and is not being addressed by the civilian research community. Further, a tri-service S&T program needs to coordinate the expertise resident across the service laboratories, so that it is effectively focused on this difficult problem.

8.2.1 Cloud Descriptions

A cloud is a visible mass of minute water droplets or ice particles suspended in the atmosphere. Fog is basically a cloud that reaches the surface of the Earth and is a direct expression of the physical processes that are taking place in the atmosphere.

In order to produce a cloud, three conditions must be met: sufficient moisture must be present, some sort of lifting (or cooling) mechanism, and condensation or sublimation nuclei to initiate the process. There have been international agreements on cloud classification, a convention accepted by most countries around the world. The importance of an international classification of clouds cannot be overestimated, since it tends to make cloud observations standard throughout the world.

Clouds have been divided into etageres, genera, species, and varieties. This classification system is based primarily on the mechanisms that produce clouds. Although clouds are continually in a process of development and dissipation, many have distinctive features that make the classification possible. Genera clouds are:

- CIRRUS (CI)—Thin feather-like clouds
- CIRROCUMULUS (CC)—Thin cotton or flake-like clouds
- CIRROSTRATUS (CS)—Very thin, sheet cloud
- ALTOCUMULUS (AC)—Sheepback-like clouds
- ALTOSTRATUS (AS)—Highly uniform sheet cloud
- NIMBOSTRATUS (NS)—Dark, threatening rain cloud
- STRATOCUMULUS (SC)—Globular masses or rolls
- STRATUS (ST)—Low uniform sheet cloud
- CUMULUS (CU)—Dense dome-shaped puffy looking clouds
- CUMULONIMBUS (CB)—Cauliflower towering clouds with cirrus veils on the top

For further information on clouds and their database and how it must be coupled with the meteorologist's ability to predict cloud coverage and location necessary to employ "Tactical" High-Energy Laser, readers should go to Reference [31] of this chapter.

8.2.2 Absorption and Scattering of Laser Beam by Gases and Solids

Earlier in this chapter, we discussed the absorption and scattering of light by both gases and solids. While the atmosphere is composed primarily of gases, solids are present, too, in the form of suspended particulate matter (water droplets and dust or aerosols). Both contribute their part to energy losses from a propagating laser beam. In our discussion of fundamentals, we concentrated our attention on the interaction of a single photon of light with a single molecule of gas. Our task now is to extend that analysis to the case where many photons of light encounter many molecules of gas, as well as small, suspended particles.

Molecules: When a photon encounters a molecule of gas, it may be absorbed or scattered. The probability of this happening is expressed in terms of the *cross*

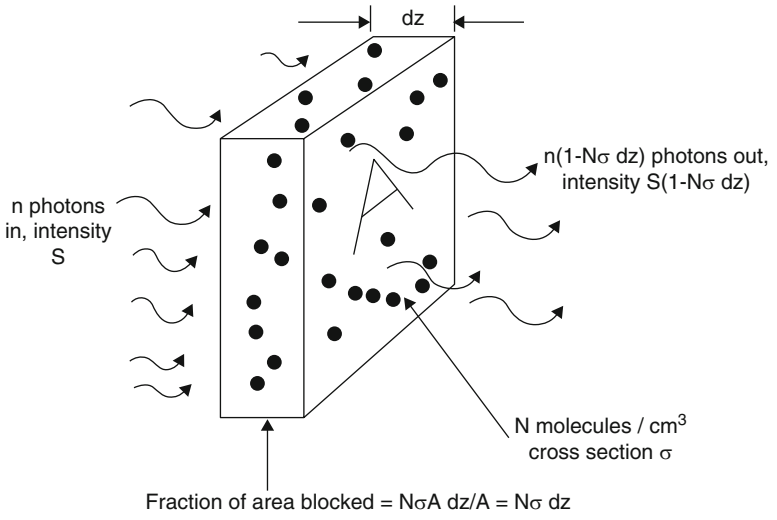


Fig. 8.1 Scattering and absorption cross section

section σ[1] for such an event to occur [32]. This concept is illustrated in Fig. 8.1. Imagine that a laser beam of area A is propagating through a thickness dz of the atmosphere in which there are N molecules per cubic centimeter. The total number of molecules that photons within the beam will encounter is NAdz. If each of these molecules has an effective "size" or cross section σ, the area blocked off by the molecules will be $a = \sigma NA$dz. Therefore, the probability that a photon will collide with a molecule and be lost from the beam through absorption or scattering is the ratio of the area blocked off to the total area, $a/A = N\sigma$dz.

This means that if n photons enter the region shown in the figure, $nN\sigma$dz will be lost from the beam. Since the beam intensity, S, is proportional to the number of photons n, it follows that S decreases by an amount $\mathrm{d}S = -SN\sigma\mathrm{d}z$ in propagating a distance dz.

The equation $\mathrm{d}S = -SN\sigma\mathrm{d}z$ is well known in mathematics. Its solution for the intensity $S(z)$ which a beam whose original intensity was $S(z)$ will have after propagating a distance z is $S(z) = S(0)e^{-N\sigma z}$. This is known as *Bouguer's law* or *Lambert's law*. This law simply states that as light propagates through the atmosphere (or any substance, for that matter), its intensity decreases exponentially over the distance traveled. The quantity $N\sigma$ is traditionally denoted K and is called the *attenuation coefficient*. This distance over which a beam's intensity will decrease by a factor of $1/e$ (about 1/3) is $1/K$, called the *absorption length*. The product $Kz = N\sigma Z$ is known as the *optical depth* and is a measure of the effective thickness,

[1]The treatment of binary (one-on-one) interactions in terms of cross sections is common in physics and will be used throughout this book. A good discussion of the cross-sectional concept can be found in Chapter II of Reference [32].

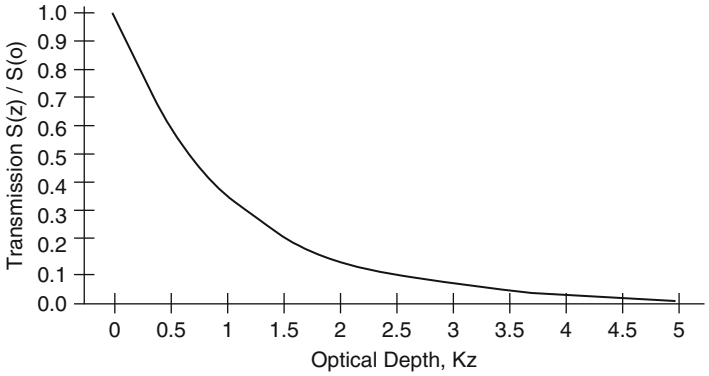


Fig. 8.2 Transmission vs. optical depth

from the standpoint of absorption, of the medium through which the light has traveled.

Figure 8.2 is a plot of $S(z)/S(0)$, the fraction of a beam's intensity transmitted over a range z, as a function of optical depth Kz. You can see from this figure that for ranges z much greater than $1/K$, larger amounts of energy will be lost from the beam. Clearly, we must choose the parameters of a laser so that K is as small as possible, and the effective propagation range, $1/K$, as larger as possible.

Our derivation of the absorption law looked at the probability of a photon interacting with a single type of molecule, with a cross section σ. Within the atmosphere, there are many types (species) of molecules present (N_2, O_2, CO_2, etc.). The probability of interaction with one is independent of the probability of interaction with another. That is, photons lost through the interaction with one type of molecule may be added to those lost through the interaction with another. As a result, the attenuation coefficients attributable to each type of molecule may simply be added: $K = K(N_2) + K(O_2) + K(CO_2) + \ldots (\text{etc.})$. Furthermore, the attenuation coefficient due to each molecular type is in turn comprised of two parts, one for the absorption of photons by that type and one for their scattering: $K(CO_2) + K$ (absorption by CO_2) $+ K$ (scattering by CO_2) and so on.

Clearly, attenuation in the atmosphere can be quite complex, with a variety of terms contributed from different molecular species, whose relative abundance and importance might change with latitude, longitude, relative humidity, and other climatic factors. Each of these terms may have a different dependence upon the laser wavelength, since a given wavelength may not be absorbed by one type of molecule and yet be strongly absorbed by another.

Therefore, it should not be surprising that a considerable body of literature has developed in this area, ranging from detailed studies of the absorption by a single molecule to gross measurements of how much light penetrates the atmosphere as a function of frequency under given climactic conditions [33]. We can only scratch

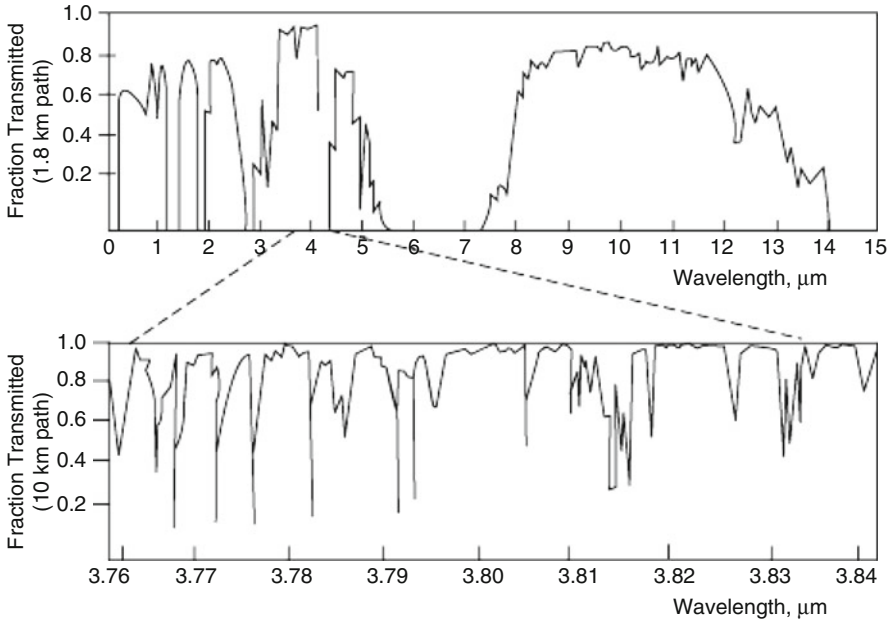


Fig. 8.3 Atmospheric attenuation vs. wavelength

the surface and provide a general feeling for atmospheric absorption and scattering.[2]

Figure 3.24 is an overview of atmospheric attenuation over a broad range of wavelengths [33]. This figure shows some of the broad windows for the propagation of laser light. However, there is considerable fine structure which the scale of this figure does not reveal. This is shown in the bottom portion of Fig. 8.3, which is an expanded view of one narrow region in the upper half.

From Fig. 8.3, you can see that even within what appears to be a propagation window, there may be narrow absorption bands at specific frequencies. Therefore, the choice of laser wavelength can be critical for propagation. For example, recent measurements of the output frequencies from a deuterium fluoride (DF) chemical laser have resulted in a changing one wavelength from 3.7886 to 3.7902 μm [33].

This change has been sufficient to alter the assessment of how much of this light would penetrate a 10-km path at sea level from about 90 % to about 50 %. This is one reason why free-electron lasers have recently received considerable interest. Unlike most lasers, whose output frequencies are fixed by the active, light-producing material in them, free-electron lasers are tunable in wavelength, offering greater flexibility in adjusting the output for efficient atmospheric propagation.

[2] A good summary of absorption as a function of frequency, adequate for zero-order analysis, can be found in Section 14, "Optical Properties of the Atmospheric" in Reference [33].

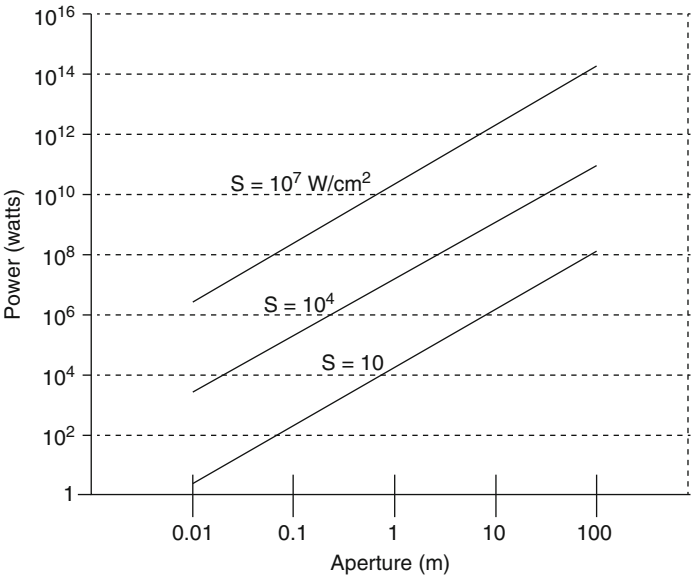


Fig. 8.4 Beam power vs. aperture and intensity

To this point, we have seen that the intensity of a laser beam will decrease with distance as $S(z) = S(0)e^{-Kz}$, where K, the attenuation coefficient, is a sum of terms representing absorption and scattering by the different species present in the atmosphere. Given K and knowing the range z required for a given application, Fig. 8.2 can be used to evaluate the resulting decrease in intensity or brightness. This decrease can then be used to modify our results for the propagation in a vacuum, allowing us to develop new criteria for target damage in atmospheric applications. For example, Fig. 8.4 says that a laser power of 10 kW is required to deliver an intensity of 104 W/cm^2 in a collimated beam with a 3-cm aperture. If the beam is propagating in the atmosphere and K is such that only 50 % of the intensity is transmitted over the range to the target, we would need to use a 20-kW laser, so that after 50 % attenuation we would have 10-kW leftover to meet the intensity requirement on the target. Alternatively, we might choose a laser with a different wavelength, for which the attenuation would be less.

A further complication arises in longer-range strategic applications. The atmospheric parameters which determine K may change over a long range, so that K is not constant, but varies with distance. This would occur, for example, in using a ground-based laser to attack the moon. As the beam goes up through the atmosphere, K, which is proportional to the density of molecules, is steadily decreasing. Eventually, the beam leaves the atmosphere and over the greater part of its range is propagating in a vacuum. We would greatly overestimate the amount of beam attenuation by using exponential absorption with a K appropriate to the atmosphere at sea level and a z equal to the range to the moon! In cases like this, we must modify our treatment of attenuation and allow K to be a function of z. If K is a

variable, dependent upon z, the solution to our original equation, $(\mathrm{d}S(z)/\mathrm{d}z) = -KS$, becomes as

$$S(z) = S(0)\exp\left[-\int_{z=0}^{z=z} K(z)\mathrm{d}z\right] \tag{8.1}$$

This "improved" version of the exponential attenuation law looks complicated, but its interpretation is straightforward. It says we must *integrate* K over the path length. In effect, we split the beam's path into many small segments. Over each, K is effectively constant, and exponential attenuation can be used. The total effect is then given by the sum of the optical depths over each small path segment. Doing this in any realistic case requires the use of a computer model which can keep track of how the distribution of molecules and their density varies with altitude and can use these data to calculate an altitude-dependent attenuation coefficient. There is, however, a simple model which is reasonably accurate, can be solved analytically, and gives a good feel for the effect of altitude dependence upon beam attenuation.

Within the lower atmosphere (0–120 km), density varies exponentially with altitude [34]. This result is seen experimentally and may be derived using statistical mechanics. It assumes that temperature is roughly independent of altitude and that the acceleration due to gravity is a constant. Therefore, it is most accurate near the surface of the Earth. Since this is where absorption is the greatest, the "exponential atmosphere" is often adequate for "zero-order" analysis. That is, the density of molecules $N(h)$ at altitude h is related to the density $N(0)$ at sea level by the relationship $N(h) = N(0)\exp(-h/h_0)$, where the constant ho is about 7 km. Since $K = \sigma N$ is also proportional to N, we can to a first approximation say that $K(h) = K(0)\exp(-h/h_0)$.

Note that since different atmospheric constituents have different molecular weights, they each fall off differently with altitude. The "7-km" value for h_0 is an average over all constituents. Species whose weight is lighter than the average will fall off less rapidly, and those whose weight is heavier than the average will fall off more rapidly. Since absorption may depend on a single species at a given wavelength, the exact scale length for absorption may differ from the nominal value in a specific application.

Suppose that we are to fire a laser into the air at some angle ϕ, as illustrated in Fig. 8.5. The beam's altitude h is related to its range z and the elevation angle ϕ through the simple geometrical relationship $h = z\sin\phi$ sin . When $\phi = 0$, for any z, and when $\phi = 90^{\circ}$, h and z are identical. Using this relationship between h and z and assuming $K(h) = K(0)\exp(-h/h_0)$,we can evaluate the optical depth to any range z. The result is shown in Fig. 8.6.[3]

Figure 8.6 is a plot of the optical depth to a range z, normalized to $K(0)h_0$, as a function of z, normalized to h_0. At $\phi = 0$, the beam is propagating horizontally, the

[3]Figure 8.6 is a plot of the expression $\int_0^z K(z)\mathrm{d}z = [K(0)h_0/\sin\phi][-\exp(-z\sin\phi/h_0)]$.

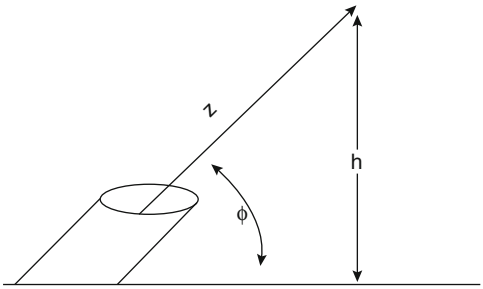


Fig. 8.5 Beam range and altitude

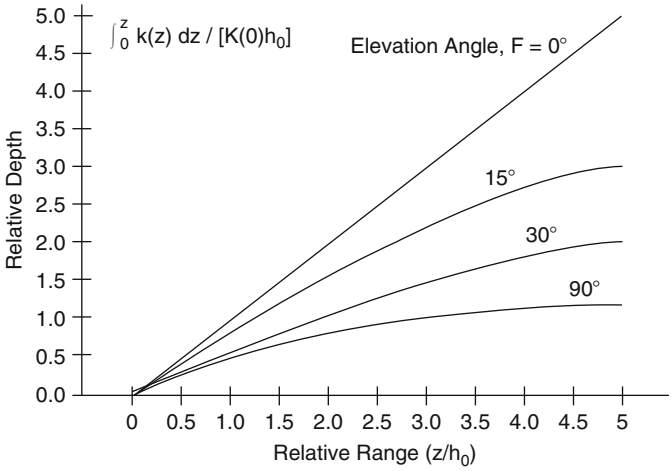


Fig. 8.6 Optical depth vs. range and elevation angle

atmospheric density is constant, and the optical depth increases linearly with distance, as we would expect. At $\phi = 90^\circ$, the beam is propagating straight up and rapidly emerges from the atmosphere, and beyond that point the optical depth no longer increases.

At intermediate angles, the beam has greater and greater lengths of the atmosphere to propagate through before it emerges from the atmosphere, and so the optical depth approaches a limiting value later and of a higher value. For ranges less than h_0, the altitude over which atmospheric density changes significantly, the optical depth is roughly independent of the elevation angle.

Figure 8.6 can be used together with our results for vacuum propagation to estimate the attenuation for long-range applications and its impact on laser requirements. For example, let us return to the case where we wish to attack the moon. If at sea level, the attenuation coefficient at the frequency of our laser is 0.1 km^{-1}, and if we are able to shoot when the moon is directly overhead $(\phi = 90^\circ)$, then by Fig. 8.6 the optical depth will be $K(0)h_0$ or 0.7.

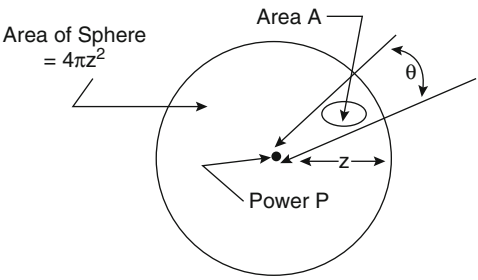


Fig. 8.7 The concept of brightness

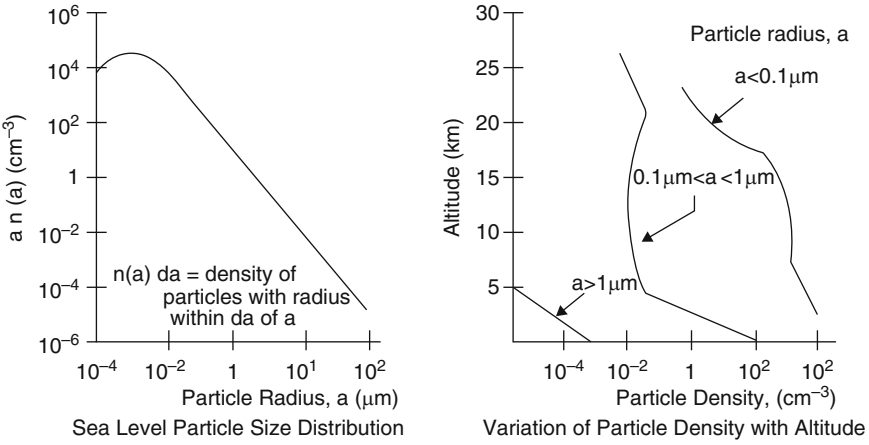


Fig. 8.8 Particle size distribution and variation with altitude

We can then use Fig. 8.2 to see that for this optical depth, laser intensity and brightness will be about 0.6 of what they would have been in a vacuum. Therefore, the brightness requirements necessary to place a given intensity on the target, obtained from Fig. 8.7, must be increased by a factor of 1/0.6 or about 1.7.

Small Particles (Aerosols): To this point, we've looked at the attenuation of a laser beam due to the gases (molecules) which comprise the majority of the atmosphere. We must next consider the effect of small solid or liquid aerosols which are invariably suspended in the atmosphere, especially near the surface. Figure 8.8, for example, shows the number density of suspended particles as a function of particle radius at sea level, along with the way in which the density of particles in different size ranges varies with altitude.[4] This figure must, of course, be considered somewhat notional, since the actual particle size distribution can vary greatly, depending on the local climate and wind conditions. A commonly used

[4]Figure 8.8 is adapted from figures found in C. E. Junge, *Air Chemistry and Radioactivity* (New York: Academic Press, 1963), and n J. E. Manson's article in S.L. Valley (ed), *Handbook of Geophysics and Space Environments* (Hanscom AFB: Air Force Cambridge Research Laboratories, 1965).

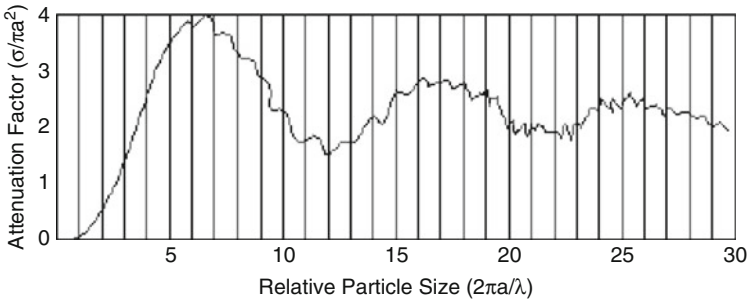


Fig. 8.9 Attenuation factor due to aerosols in Mie's theory [36]

mathematical expression for the density of particles of size r is $r = ar^{\alpha}\exp(-br^{\gamma})$. The constants a, b, α, and γ will vary depending on climate and other conditions. Representative values may be found in Sect. 3.14 [35]. A number of important things are apparent from Fig. 8.8. First, particles in excess of 1 μm are quite rare and largely confined to regions near the surface of the Earth. Second, the range of particle sizes is comparable to the wavelengths of lasers operating from the visible to far infrared (0.4–10 μm). The absorption and scattering of light by solid particles become quite complex to analyze when the particle size is comparable to the wavelength of the light.

The relevant theory is known as *Mie scattering theory* for its developer, a German meteorologist [36]. The development of this theory is beyond the scope of this book, but its essence is summarized in Fig. 8.9.

Figure 8.9 shows how the actual attenuation cross section for a dielectric aerosol (in this case water) compares to its physical size, πa^2, as a function of $2\pi a/\lambda$ where a is the aerosol radius and λ the wavelength of the light. There are various dips and bumps in the cross section, reflecting resonances between particle size and light wavelength, but for the most part, σ is on the order of $2\pi a^2$, especially when a is much larger than λ. Somewhat crudely, you might think that each particle contributes twice its physical cross section to light attenuation because it can contribute to attenuation in two ways—through absorption and scattering.

Curves similar to Fig. 8.9 are available in the literature for a variety of different particle types, both dielectric and metallic[5]. As a general rule, they exhibit behavior similar to that shown in Figure 8.9—σ falls to zero as $2\pi a/\lambda$ goes to zero, but for $2\pi a/\lambda > 1$, σ is in the neighborhood of $2\pi a^2$. Given Fig. 8.8, which suggests that the majority of aerosols are of a size less than 1 μm, we can conclude that the effect of aerosols on light attenuation will be greater for visible lasers ($\lambda = 0.4 - 0.7\,\mu\text{m}$) than for those operating in the infrared ($\lambda = 1 - 10\,\mu\text{m}$). The relative particle size, $2\pi a/\lambda$, is shown in Fig. 8.10 as a function of wavelength and particle size. This figure may be used with Fig. 8.9 to estimate the contribution of aerosols to light

[5] A number of representative curves, along with references to the original literature, can be found in Born and Wolf Reference [36].

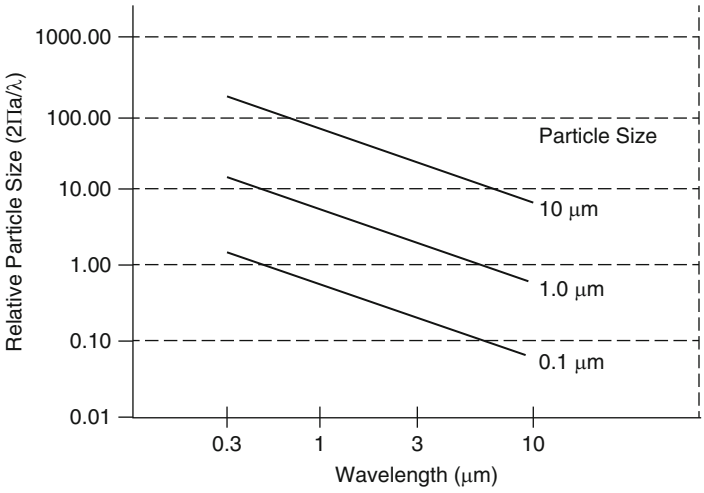


Fig. 8.10 $2\pi a/\lambda$ vs. wavelength and particle size

attenuation. However, it must be emphasized that under realistic circumstances the density and type of aerosols can vary greatly from day to day.

Thus, the operational use of lasers will require that site-specific surveys be made and that beam brightness be increased to enable meeting damage criteria under "worst-case" conditions.

Just as the attenuation coefficient for a mixture of molecules may be obtained as the sum of the coefficients contributed by each molecule present, the total attenuation coefficient when aerosols are suspended in the atmosphere is the sum of contributions from the molecules and the aerosols taken separately. Molecules and aerosols can contribute to attenuation in roughly equal amounts, but for quite different reasons. The attenuation coefficient K is σN, where σ is the attenuation cross section and N the density of the attenuator. Molecules have a very small attenuation cross section ($10^{-25} - 10^{-26}\,\mathrm{cm}^2$), but their density is quite large ($3 \times 10^{19}/\mathrm{cm}^3$ at sea level). By contrast, particles can have a very large cross section $\left(\approx 3 \times 10^{19}/\mathrm{cm}^8\mathrm{cm}^2\right)$, but their density is quite low ($1/\mathrm{cm}^3$ or less). Further details about the variation of atmospheric density with altitude and its effects on energy absorption and scattering can be found in the book by P. E. Nielsen *Effects of Directed Energy Weapons*, published in 1994, and the PDF format of it can be found on the Internet to download. All the above portion of this book (Sect. 8.2.1) was copied from his book, and we encourage readers to refer to this title. There is more information such as the index of refraction variations of air, and its impact on turbulent and coherency beam also has been discussed by him as well. He also shows a good section on adaptive optics that makes use of the fact that if we know what the atmosphere is like along the beam path, it's possible to send out the distorted beam in such a way that turbulence will in fact straighten it out.

To discuss further the *aerosol absorption and nonlinear effects of the radiation propagation in aerosols*, we should note that the aerosol absorption is a more regular function of the wavelength than molecular absorption. This statement is illustrated in Fig. 8.8 which shows the coefficients of attenuation and absorption spectra calculated in footnote 4 for conventional models of the continental and sea aerosols.

The spectra are normalized by the attenuation coefficient α_0 at a wavelength of $\lambda = 0.55\,\mu\text{m}$. The coefficient varies over a wide range with the weather. In a transparent atmosphere with a visibility range of $S > 25$ km, the coefficient is $\alpha_0 < 0.15\ \text{km}^{-1}$. Coefficients of absorption by the continental aerosol at the wavelengths of high-power lasers are $\alpha_{\text{abs}} < 0.002\,\text{km}^{-1}$ for a DF laser ($\lambda = 3.75\,\mu\text{m}$), $\alpha_{\text{abs}} < 0.003\ \text{km}^{-1}$ for a CO laser ($\lambda = 5.05\,\mu\text{m}$), and $\alpha_{\text{abs}} < 0.006\,\text{km}^{-1}$ for a CO_2 laser ($\lambda = 10.6\,\mu\text{m}$), which is approximately one-tenth of the molecular absorption coefficients. Meanwhile, in the ranges $\lambda = 0.69\,\mu\text{m}$ and $\lambda = 1.06\,\mu\text{m}$, the aerosol absorption is comparable with the molecular one even in the case of good visibility and should be taken into account in the calculation of thermal nonlinearities. The average temperature and refractive index variations depend, as in the case of the molecular absorption, only on the aerosol absorption coefficient.

Unlike the molecular absorption-induced heating, the air heating resulting from absorption by aerosol particles is nonuniform. There are thermal halos, i.e., zones of an elevated temperature, around the particles. Heat halos lead to higher radiation scattering. The effect was predicted by Askaryano [37] and experimentally observed in References [38] and [39]. The scattering-induced variation of the radiation attenuation coefficient is found from Reference [39] (Fig. 8.11):

$$\alpha_{\text{att}} = \alpha_{\text{att}}^0 \left[1 + \left(\alpha_{\text{abs}}/\alpha_{\text{att}}^0\right) f^2(I_0, t)\right] \tag{8.2}$$

$$f(I_0, t) = \frac{\mathrm{d}n}{\mathrm{d}T} \frac{I_0}{\rho_0 c_{\text{rel}}} \sqrt{\frac{\sigma_{\text{abs}} t_1}{\lambda^2 \chi}} \tag{8.3}$$

$$f_I = 9 \int_{x=0}^{x=\infty} x \left\{ \frac{\sin x - x\cos x}{x^3} \int_{\tau'=0}^{\tau'=\tau} \frac{I(\tau' t_0)}{I_0} \exp\left[-x^2(\tau - \tau')\right] \mathrm{d}\tau' \right\}^2 \mathrm{d}x \tag{8.4}$$

where α_{att} is the scattering-induced attenuation coefficient for a low beam intensity; σ_{abs} is the aerosol absorption coefficient; cab is the particle absorption cross section; I is the radiation intensity density; n, ρ_0, c_{rel}, and χ are the refractive index, density, heat capacity, and thermal conductivity of the air, respectively; λ is the wavelength of the light radiation; $\tau = t/t_0$; $t_0 = a^2/\chi$; and a is the radius of the absorbing particle. At times $t \gg t_0$, and with the relationship $I(t)$ of the form $I(t) = I_0$ for $t \gg t_0$ true, time t_1, determined from Eq. 8.4, is directly proportional to the momentum time t, i.e., $t_1 = t\ln 2$. Substituting air parameters in Eq. 8.4, we have

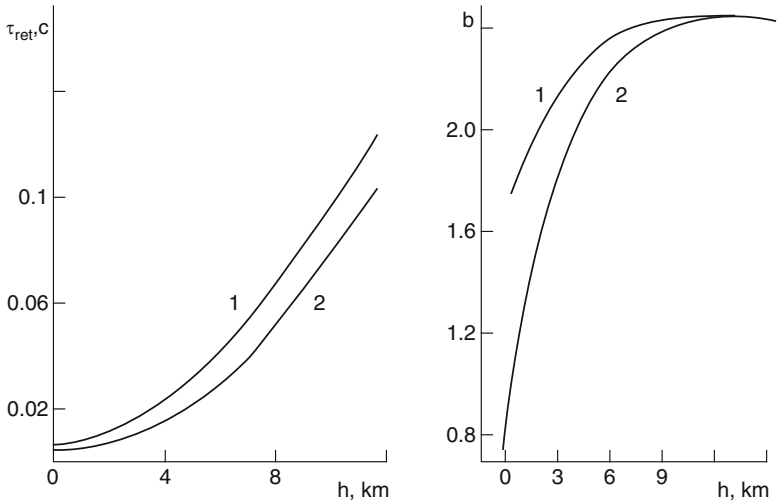


Fig. 8.11 The transition time τ_{rel} of the absorbed energy to thermal one and the parameter δ determining the sign of the air temperature variation as a function of altitude (h) [17]. With $\delta < 1$, the air is heated; with $\delta > 1$ it is cooled: 1–30 % H_2O, 2–100 % H_2O

$$f(I_0, t) = 10^{-3} I_0 \sqrt{\left(\sigma_{abs}\left(\lambda^2\right) \cdot 4t\right)} \tag{8.5}$$

where I_0 is given in W/cm^2 and t in s.

Evaporation of water drops in the laser radiation field has recently been in the focus of research into possible clearing of clouds and fog (Fig. 8.12).

The radiation power necessary for drop evaporation can be estimated for the wind velocity, v_0, as

$$P = (\alpha_{abs}/\alpha_{abs})(\pi d v_0/4) q l L_v \tag{8.6}$$

where d is the beam diameter, l is the length of the layer of the cloud to be evaporated, q is the water content, and L_v is the latent heat of evaporation. For $d = 1\,\text{m}$, $v_0 = 5\text{m/s}$, $(\alpha_{abs}/\alpha_{abs}) = 0.1$, $l = 100\,\text{m}$, and $q = 1\ \text{g/m}^3$, the power $P \approx 10^7\text{W}$.

It follows from the above estimate that high power is required for cloud clearing, i.e., laser beams could transmit high powers effectively only in a clear atmosphere. For further details of such analysis, please refer to a report by V.V. Vorob'ev [40].

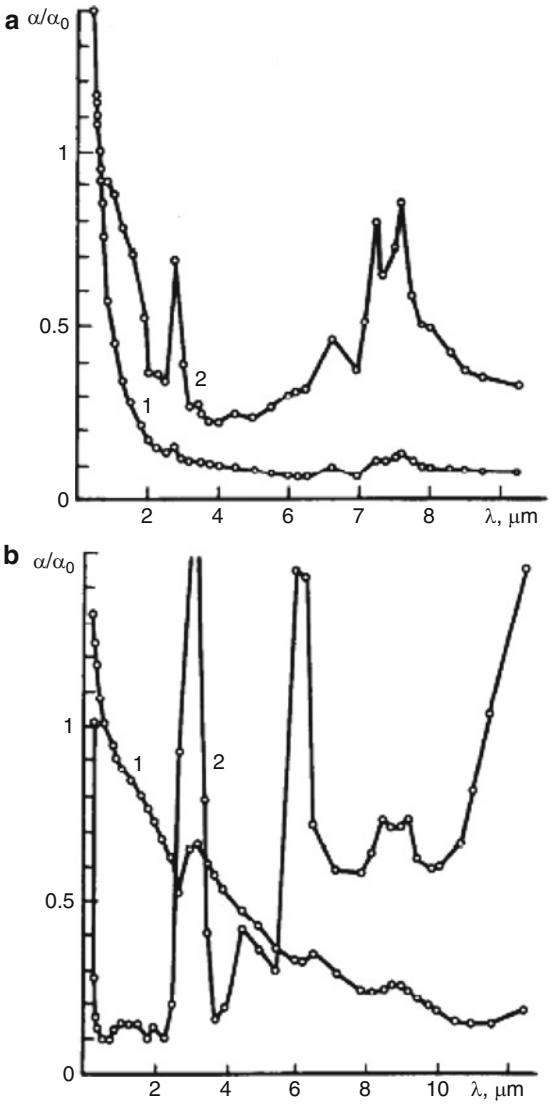


Fig. 8.12 Spectra of the coefficients of attenuation α_{att} and absorption α_{abs} by the atmospheric aerosol. Conventional models of the continental and sea aerosols, respectively (α_0 is the attenuation coefficient on the wavelength $\lambda = 0.55$ pm). The values of $\alpha_{abs}/\alpha_{abs}$ for the sea aerosol on the wavelengths $\lambda = 3.00$ and 3.20 μm are equal to 0.33 and 0.22, respectively

8.3 Laser and Thermal Blooming Effects

To this point, everything we have considered has been independent of the laser's intensity. Absorption, for example, reduces the intensity at a given range z by a fraction, $S(z)/S(0) = e^{-Kz}$, which is independent of $S(0)$. This means that the transmitted intensity, $S(z)$, is directly proportional to the intensity out of the laser, $S(0)$. If we were to plot $S(z)$ as a function of the $S(0)$, the plot would be a straight

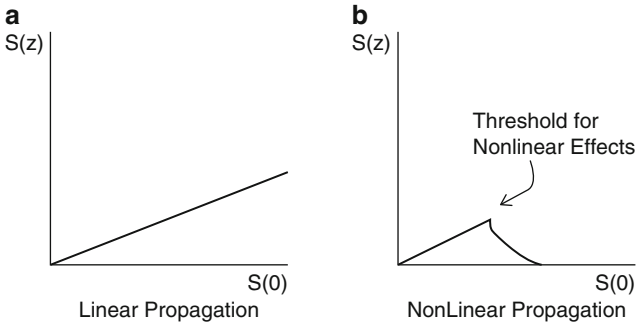


Fig. 8.13 Linear and nonlinear propagation effects

line, as shown in Fig. 8.13a. For this reason, the propagation effects we've looked at so far are known as linear effects.

As intensity increases, it is usually found that at some point the relationship between $S(z)$ and $S(0)$ is no longer linear—there is a sudden shift in behavior, like that shown in Fig. 8.13b.

This typically occurs when $S(0)$ exceeds some threshold value, and beyond that threshold the relationship between $S(z)$ and $S(0)$ can be quite complex and not at all linear, so that *nonlinear* propagation effects are said to have occurred. The physical reason for nonlinear effects is that when intensities are strong enough, the beam actually modifies the environment through which it propagates in such a way that its physical characteristics are altered. For example, a very intense beam might vaporize the aerosols in its path and suffer less degradation due to aerosol absorption and scattering than a lower intensity beam. Or it might ionize the atmosphere in its path, increasing absorption to the point where propagation ceases. Unfortunately, most nonlinear effects degrade, rather than enhance the intensity on the target. In this section, we will consider nonlinear effects which affect both propagation (thermal blooming and bending) and attenuation (stimulated scattering, breakdown, and absorption waves). Our emphasis will be on those effects which are of concern from the standpoint of beam propagation in weapon applications. There are many nonlinear effects not considered here which may be important at the intensities and powers appropriate to other applications, such as laser fusion, or which are of interest from a scientific standpoint for insight into the structure of matter.

One of the first nonlinear phenomena recognized as likely to affect the propagation of a high-power laser, thermal blooming results from the energy which a laser deposits in the air through which it propagates. The beam loses energy as a result of absorption. This energy is deposited within the beam path, where it causes a temperature rise in the air. This temperature rise modifies the air's density, alters its index of refraction, and can severely affect the beam's propagation. The sequence of events which results in thermal blooming is illustrated in Fig. 8.14.

Figure 8.14a shows the intensity profile of a typical laser beam viewed end on: higher in the center than at the edges, where the intensity falls to zero.

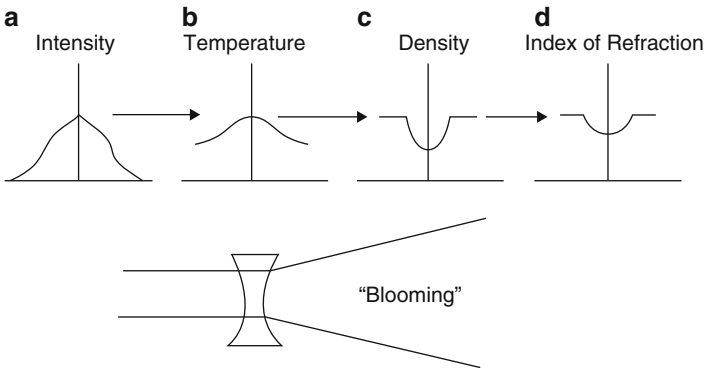


Fig. 8.14 The physics of thermal blooming

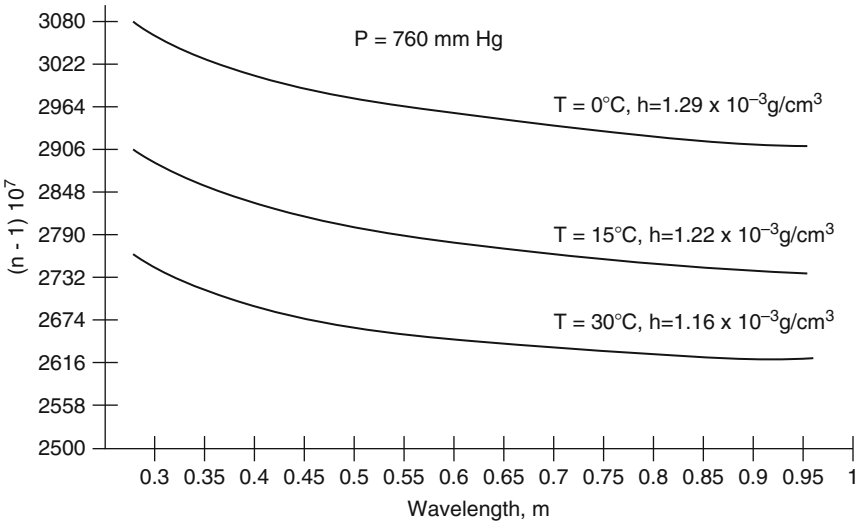


Fig. 8.15 Index of refraction of air vs. temperature and wavelength

The temperature of the air through which the beam propagates will exhibit a similar profile, as suggested by Fig. 8.14b. This is because absorption of energy by the air is a linear phenomenon—as intensity goes up, the amount of energy absorbed, being a constant fraction of the incident intensity, goes up as well.

The absorbed energy manifests itself as an increase in temperature. But hot air is less dense than cold air—at constant (atmospheric) pressure, an increase in temperature implies a decrease in density. Thus, the density profile in the air through which the beam propagates assumes a form inverse to the intensity profile, as shown in Fig. 8.14c. The implication of this is that the index of refraction of the air through which the beam propagates, shown in Fig. 8.14d, mirrors the density profile, since $(n-1)$ is proportional to density (see Fig. 8.15).

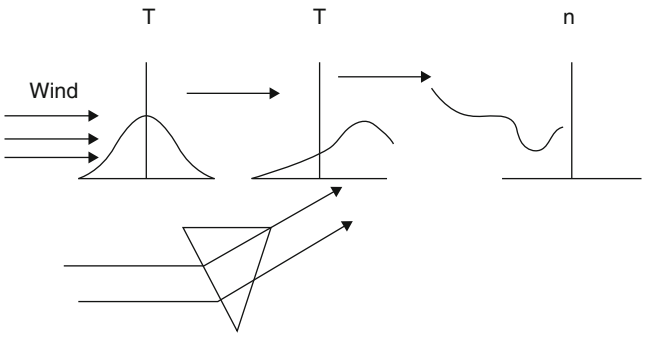


Fig. 8.16 The physics of thermal bending

This sequence of events has the effect of introducing within the volume of the beam what amounts to a diverging lens, with a greater optical density at the edges than at the center. As shown at the bottom of Fig. 8.14, this causes the beam to *bloom* or diverge, at a rate greater than would otherwise be expected.

In practical scenarios, wind sweeps across the beam—either a naturally occurring wind or one which results from the relative motion of the beam and atmosphere as the beam slews to keep itself on the target. Such a crosswind causes the beam to *bend* as well as bloom. The physical reason for this is illustrated in Fig. 8.16. In this figure, we begin with the temperature profile which produced thermal blooming. However, wind is blowing across the beam and introduces cold air. As a result, the upwind portion of the temperature profile becomes cooler and the downwind portion warmer. In effect, the wind tries to push the hot air downstream. As a result, the index of refraction profile assumes a shape like that shown, and the beam sees what looks like a wedge inserted into it.

This wedge causes the beam to bend into the wind, as indicated. Another way of looking at it might be to think that the wind displaces the diverging lens of thermal blooming, so the beam sees only one-half of that lens, and bends in a single direction.

Looking at Figs. 8.14 and 8.16, you can easily imagine that the analysis of thermal blooming and bending will be quite difficult, involving gas flow, laser heating, and the temperature and density dependence of the index of refraction of air. In any realistic scenario, wind velocity will vary along the beam path, and beam blooming and bending will occur simultaneously. The resulting distortions of the beam's intensity profile can be quite complex, as shown in Fig. 8.17, which compares a beam's intensity on the target in the presence of these effects with what it would have been in their absence. Our goal will be to determine the thresholds for thermal blooming, the magnitude of the effect, and the potential for dealing with it through techniques such as adaptive optics (see Sect. 8.4) in those cases where target range and damage criteria prevent the operation below the threshold. We must first recognize that there is a pulse width threshold for thermal blooming. Even though the atmosphere absorbs energy and its temperature begins

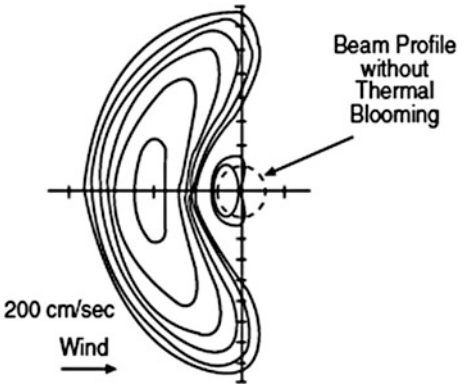


Fig. 8.17 Beam profile with thermal blooming and bending [41]

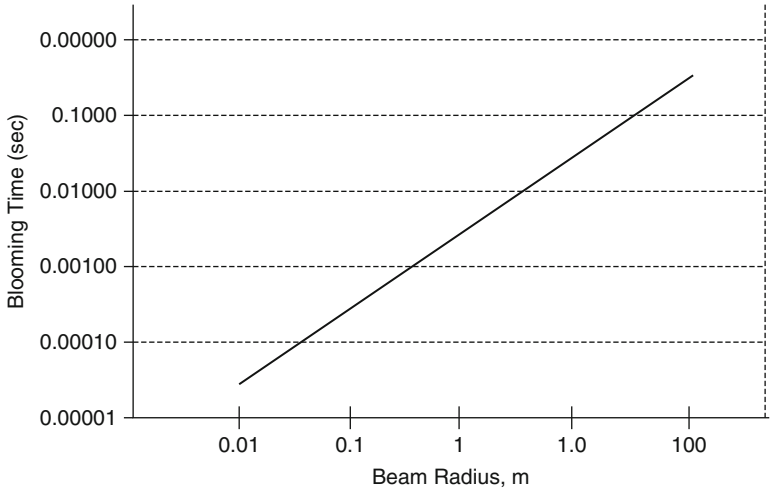


Fig. 8.18 Time for the onset of thermal blooming vs. beam radius

to rise almost immediately, some finite time is required for the heated air to expand and move out of the beam, creating the density "hole" shown in Fig. 8.14c. The characteristic velocity at which disturbances propagate in air is the speed of sound, $a_c = 3 \times 10^4$ cm/s. Therefore, the time for air to move out of a beam of radius w and for blooming to begin is approximately w/a_c. Figure 8.18 is a plot of the time for thermal blooming to develop as a function of beam radius.

You can see from Fig. 8.18 that thermal blooming is not likely for pulsed lasers, where the goal is to place all the energy on the target in short time scales—10^{-5} s or less, for example. On the other hand, if seconds of interaction time with the target is required, there is a potential for thermal blooming to be a problem, even for strategic applications, where the beam radius may be relatively large. Therefore,

we must consider next the magnitude of the effect, to see if thermal blooming will pose a serious threat to mission accomplishment should it occur.

The quantitative analysis of all nonlinear phenomena is quite complex and difficult, since it depends on a beam intensity which is itself changing in response as the interaction proceeds. Predicting how the spot size and intensity vary with time on the target for a beam of arbitrary intensity distribution therefore requires a computer program in which all these effects are modeled [42]. Considerable insights can be gained, however, by looking at simplified examples for which exact solutions exist. One case which has been extensively studied is that of a beam in a uniform crosswind of velocity v, having an intensity profile [43] which varies with the radius as $S(r) = S_0 \exp(-2r^2/w^2)$. Such a beam can be characterized through a *thermal distortion factor*, N_t, which is given by [44] as

$$N_t = -\frac{(dn/dT)}{n\rho c_p} \times \frac{KSz^2}{vw} \tag{8.7}$$

The first factor in the expression for N_t contains parameters related to the gas through which the laser is propagating, and the second contains parameters related to the laser and scenario in which it is employed. The individual terms have the following interpretation:

- dn/dT is the slope of a curve of the index of refraction, n, as a function of temperature, T. The greater the dependence of n on T, the more pronounced will be the lens or wedge introduced as the beam heats the air.
- c_p is the heat capacity of the air (J/g K), and ρ its density (g/cm^3). Their product, ρc_p, is the number of joules of energy which must be absorbed to heat a cubic centimeter of air by 1°.
- K is the absorption coefficient of the air (cm^{-1}), and S the laser intensity (W/cm^2). Their product, KS, is the number of joules being deposited in a cubic centimeter of air each second.
- z is the range to the target, w the beam radius, and v the wind velocity. N_t increases as z goes up, because the thermal lens has a longer distance over which to act. It decreases as v and w increase. A stronger wind will cool the beam volume, perhaps even blowing the heated air out of it. Since the intensity changes the most near the edge of the beam, a larger w reduces the relative importance of these edge effects and the blooming or bending which results from them.

As N_t increases, the beam becomes more and more distorted, and its intensity falls off as shown in Fig. 8.19. Since N_t is proportional to the beam intensity S, it's not possible at large distortion numbers to overcome the effect of thermal blooming by increasing S. For example, if N_t is 10, the intensity on the target will be about 0.1 of what it would have been in the absence of blooming. If we try to compensate for this by increasing S by a factor of 10, we'll increase N_t to 100, since it's proportional to S. But at an N_t of 100, the relative intensity is 0.001, and the net effect of

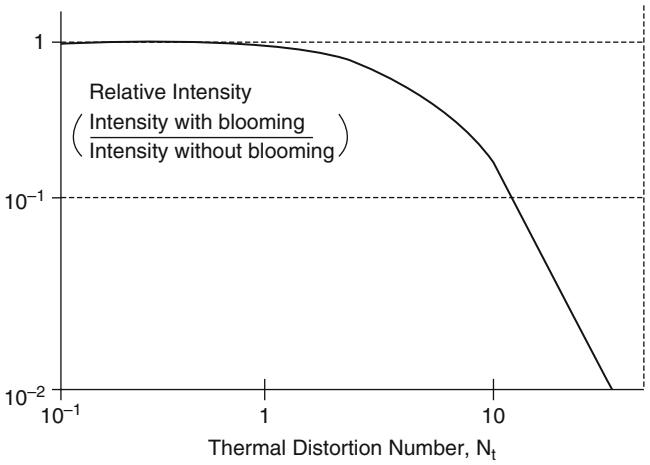


Fig. 8.19 Relative intensity vs. distortion number

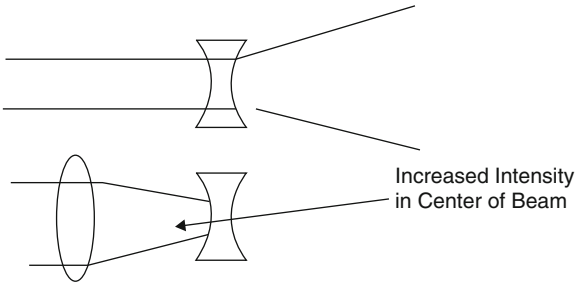


Fig. 8.20 An instability in thermal blooming with adaptive optics

increasing S by a factor of 10 will have been to reduce the intensity on the target! This illustrates one of the more unpleasant features of nonlinear effects. Actions taken to correct them can have an effect opposite to that intended, because of the many feedback loops which affect how the system responds to its inputs.

It is interesting to note as another example that an instability may occur when adaptive optics (see Sect. 8.4) is employed in the presence of thermal blooming. The reason for this is sketched in Fig. 8.20.

In the upper portion of the figure, we see what an adaptive optic system perceives as thermal blooming begins—that a diverging lens has been inserted into the beam path. As a response, the system tries to do what is shown in the bottom portion of the figure—send a converging beam into the lens so that the diverging nature of the lens will only serve to straighten the beam out. Unfortunately, this serves to increase the intensity in the center of the beam, increase the temperature in this region, and aggravate the diverging lens effect, leading to further focusing, further divergence, etc. Thus the algorithm responsible for adaptive optics must be capable of adapting to and compensating for nonlinear phenomena such as thermal blooming, as well as linear phenomena such as turbulence. This can create both hardware

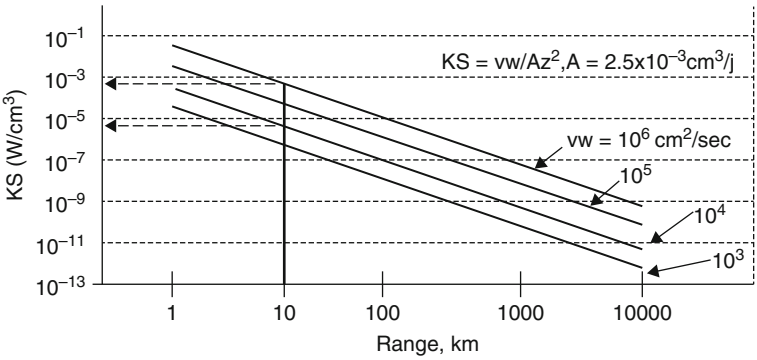


Fig. 8.21 Parameter tradeoffs to prevent thermal blooming

and software challenges in dealing with the two phenomena together. Developing adaptive optic schemes capable of handling both thermal blooming and turbulence is an ongoing area for research [43].

If we want to avoid the complexities of dealing with thermal blooming, Fig. 8.19 tells us that the distortion number must be of order unity or less. What are the implications from the standpoint of beam design? Figure 8.21 shows the relationships among intensity, range, velocity, and beam radius subject to the constraint that $N_t = 1$. The curve may be used in the following way.

Suppose we wish to engage a target at a range of 10 km with a beam whose radius is 0.5 m (50 cm) and anticipate that the crosswind will be 5 mi/h (about 200 cm/s). Then the product vw is 10^4 cm^2/s. As the lines drawn on the figure show, the product KS must therefore be less than or equal to about 3×10^{-6}W/cm^3 if the distortion number is to be kept less than 1. If we know that the beam intensity needs to be 104 W/cm^2 in order to meet our damage criteria, this implies that the absorption coefficient K must be less than 3×10^{-10} cm^{-1} (3×10^{-5} km^{-1}) for no thermal blooming or bending to occur. Since absorption coefficients within propagation "windows" are more like $10^{-3} - 10^{-2}$ km^{-1}, you can see that it's very unlikely that we can accomplish this mission without thermal blooming [43]. On the other hand, if the target is an aircraft moving at 500 mi/h, vw will be increased to 10^6 cm^2/s and KS to 3×10^{-4}W/cm^3. Under these circumstances, a 10^4W/cm^2 laser can do the job if K is less than 3×10^{-3} km^{-1}, a more reasonable value.

You can play with Fig. 8.21 and look at the likelihood of avoiding thermal blooming under a variety of scenarios. In realistic scenarios, of course, the wind is unlikely to be constant over the whole path, and the beam profile may not be Gaussian. An effective distortion number can nevertheless be calculated by integrating over factors which change along the beam path and over the beam front, much as we did in integrating $K(z)\,dz$ to obtain the optical depth when K was not a constant.

The procedure is described in Gebhardt (note 31). The relative intensity (Fig. 8.19) varies with the distortion number as in Fig. 3.42 for a surprisingly

broad range of experimental conditions. Such an examination will convince you that there are far more circumstances where thermal blooming needs to be considered, than where it does not. The only sure-fire way around the problem is to shorten the interaction time to the point where blooming cannot develop (Fig. 8.18). This solution can bring problems of its own with it, however, since at short pulse widths, the intensity necessary for the damage can become quite high and may exceed the threshold for other nonlinear effects, such as stimulated scattering and air breakdown.

In summary, propagation of light beams through the atmosphere is accompanied by an energy loss by molecular absorption and by scattering on aerosols and molecules. For light power laser, the air heating due to absorption induces along the propagation path a reduction effect of the refractive index in the beam axis, whose effect is to spread the beam and to noticeably reduce in radiance relative to its expected value: it is the "thermal blooming" phenomenon, which has been, for over the years, the subject of much theoretical and experimental work [17, 45].

The result of the above argument is a deflection and defocusing of the beam in a complex fashion which was discussed. The effect as we said is known as thermal blooming. Laser wavelengths for atmospheric propagation are chosen where absorption is weak; nevertheless, residual absorption together with the intense radiation can lead to blooming. The absorbed energy per unit volume which produces the heating is the product of the absorption constant, the beam intensity, and a time which can depend on beam motion through the air and heat transport due to convection (wind) or conduction. The phenomenon of thermal blooming was observed in liquids not long after the invention of the laser and was subsequently found in solids and gases. Because of the practical difficulty of full-scale atmospheric tests, much of the work has concentrated on laboratory simulation and theory.

To analyze the effects of thermal blooming, computing models and codes have been established by various authors. The experimental verification of the results of these calculations in free field over long paths presents great difficulties in view of the lack of precision in our knowledge of the various physical parameters involved in this phenomenon.

8.4 Mission Impact

It is convenient to divide cloud impacts on tactical aircraft missions into two separate categories: impacts on air-to-air missions and impacts on air-to-ground missions. It is now possible to estimate which cloud etageres might affect the prosecution of these two distinct missions. For example, high clouds would generally be of no concern to the application of a tactical laser in a close air support scenario, but might be in the en route escort of a high-value airborne asset. First, a few of the possible missions affected by low clouds are highlighted (see Table 8.1).

Table 8.1 A few of the possible missions affected by low clouds

Low cloud types	Droplet size	Possible missions affected
Stratus	100 m to 1 mm	Close air support (CAS) Offensive counter-air (OCA)
Stratocumulus	1 mm +	RECCE/CSAR, strategic attack
Nimbostratus	~1 mm +	Air interdiction mine warfare
Cumulus	10 mm to cm	CAP/TAC (A)

Table 8.2 The effects of middle clouds

Middle cloud types	Droplet size	Possible missions affected
Altostratus	10–100 m	TBM intercept
Altocumulus	1 mm +	Large A/C self-protect
Nimbostratus	~1 mm +	Small A/C self-protect
Cumulus	10 mm to cm	High-value "escort"
Cumulonimbus	10 mm to cm+	Close air support

All of the above cloud types are optically opaque in their normal state. The time of year and location will generally determine cloud content—that is, if the clouds are composed of water droplets and/or snow and ice crystals. There are, however, times when the absolute humidity present may be minimal in some stratus clouds. This condition results in an optically and physically "thin" cloud, regardless of what the surface reports may indicate. (Note: A ceiling layer of 6/10th or greater sky coverage may be reported indicating a broken to overcast sky condition, but the cloud in this case would not be optically opaque.) These "thin" ceiling layers allow the warfighter to see through the cloud and therefore would allow the possibility of employing a tactical laser. Stratocumulus clouds often present a cellular appearance with breaks in the overcast where there may be opportunities to use a tactical laser in the holes between the clouds. As mentioned previously, stratocumulus clouds are not very "thick" and do not have a large vertical extent.

The next chart Table 8.2 shows the effects middle clouds may have in the prosecution of different missions.

All of the above cloud types are optically opaque in their normal state. Again, the time of year and location will generally determine cloud content. Cumulus and cumulonimbus clouds contain by far the largest-sized water droplets (in reference to diameter) and also the greatest amount of total water vapor content. They are always optically opaque. There are, however, occasions that the absolute humidity present may be minimal in some altostratus clouds. As in the case of stratus clouds, this would result in an optically and physically "thin" cloud, regardless of what the surface reports may indicate. It is possible to see through these "thin" cloud layers and would again allow the possibility of employing a tactical laser. Altocumulus clouds present a cellular appearance with breaks in between each cell, so there may be opportunities to use a tactical laser in the holes between the clouds.

Table 8.3 The effects of high clouds

High cloud types	Droplet size	Possible missions affected
Cirrus	100 m to 1 mm	TBM intercept
Cirrostratus	100 m to 1 mm	Large A/C self-protect
Cirrocumulus	1 mm +	Small A/C self-protect
Cumulus	10 mm to cm	High-value "escort"
Cumulonimbus	10 mm to cm+	Offensive counter-air, RECCE, CAP/TAC (A)

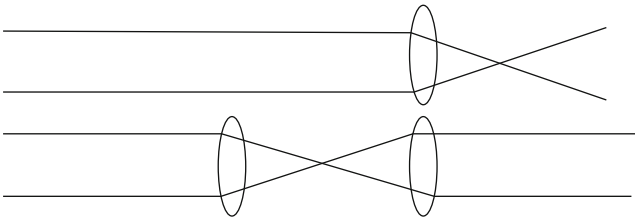


Fig. 8.22 The principle of adaptive optics

As mentioned previously, altocumulus clouds are not very "thick" and do not have a large vertical extent.

Finally, the effects of high clouds should have the least direct impact on the application of a tactically sized laser mounted on a fighter aircraft, and these are listed below (see Table 8.3).

With this brief discussion of the nature of clouds, it is now apparent that the distribution of clouds in the tactical engagement arena must be established for the warfighter to successfully employ HEL weapons. To determine the cloud distribution, the databases for historical cloud cover must be described.

8.5 Adaptive Optics

Adaptive optics makes use of the fact that if we know what the atmosphere is like along the beam path, it's possible to send out the beam distorted in such a way that turbulence will in fact straighten it out! This may seem farfetched, but the general principle is straightforward and is illustrated in Fig. 8.22. In the upper portion of the figure, a beam of light encounters a lens which might represent a cell of turbulence. This lens focuses the light, so that it diverges. In the bottom portion of the figure, a second lens has been introduced, identical to the first, and positioned at twice the focal distance from it. This lens has the effect of presenting a diverging beam to the original lens, which then proceeds to focus it back into a parallel beam. The second lens compensates for the first, so that the net result is as though the first lens were not present. The idea behind adaptive optics is to compensate for the many little "lenses" of turbulence in a similar way. All we need to do is know what's out there

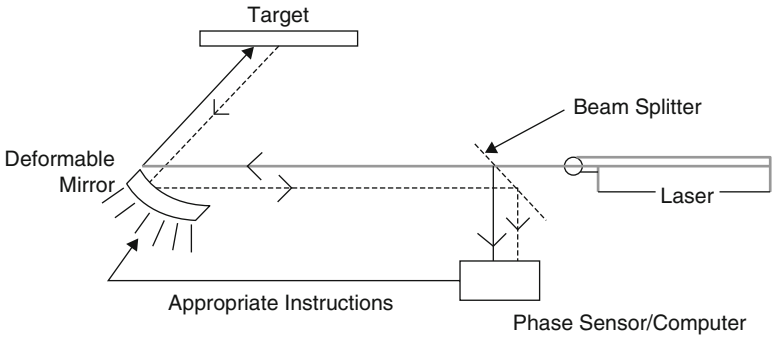


Fig. 8.23 An adaptive optic experiment

and *compensate* the beam, mixing up its initial parameters in such a way that the optical path which it traverses acts to convert the mixed-up beam into the beam we'd have in the absence of that optical path.

The practical problem in carrying out this approach is knowing what's out there so we can compensate for it, since turbulence changes from moment to moment. We need a way to obtain real-time feedback on the environment and to use that information in adjusting the beam appropriately. This is a formidable task, but not impossible, and it has been accomplished experimentally [44].

Figure 8.23 shows schematically how this has been done. The key to the technique is a *deformable mirror*—one in which small actuators move the surface up and down to distort the outgoing beam in such a way that turbulence will compensate for this distortion, resulting in a nearly diffraction-limited beam on the target. The degree of distortion required is found by examining some of the light which is reflected back from the target. This light, indicated by the dashed line in Fig. 8.23, is fed into a phase sensor, which compares the quality of the returning light with that in the outgoing beam. The difference between these is, of course, related directly to the turbulence along the beam path. A computer uses this information to provide instructions to the actuators which then deform the mirror appropriately.

There are, of course, practical considerations which may limit the ability to do this in realistic scenarios. First, the number of actuators required may be quite large. Since the beam is being broken up into segments of a size r_0, it follows that the mirror surface must be broken up into segments of area less than πr_0^2. As we have seen, propagation over large distances requires large mirror sizes—1–10 m. Since r_o is typically 5–10 cm, the mirror must be divided into something on the order of 10,000 segments, with a corresponding number of actuators. These actuators must be able to produce the necessary degree of distortion sufficiently fast to compensate for a continually shifting environment as well. Typically, this requires that any section on the mirror surface be able to move 1–10 μm in times on the order of 10^{-3} s. The time scale for mirror motion is related to the frequency with which the turbulent

environment shifts from one configuration to another, and the degree of motion is related to the wavelength of the light. See Pearson [44].

In summary, adaptive optics provides the means to maintain beam quality in the face of atmospheric turbulence. However, such methods are not applicable or completely effective in all situations. Ideal adaptive optics requires a beacon—a point source of light from the target—and this is not completely or even partially achievable in some systems or applications. Furthermore, for long slant or near-horizontal propagation paths, the integrated turbulence strength can be sufficiently strong that even the best adaptive optics cannot completely compensate for the turbulence. Thus the system performance is degraded by the atmospheric turbulence conditions and the limited capabilities of the adaptive optic system.

System performance is significantly enhanced by a capability to model and predict laser system effectiveness under specific atmospheric conditions. Forecasting and decision aids for existing conventional electro-optical systems provide a model for HEL systems. A joint program, involving the Army, Navy, and Air Force, has developed and delivered decision aids for systems where atmospheric effects are a concern, including low-light-level TV systems, passive infrared seekers, and laser-guided munitions.

The Air Force Research Laboratory leads the tri-service team that developed the Target Acquisition Weather Software (TAWS) currently in operational use by the Air Force and Navy weather support personnel in both training and strike mission planning. For IR seekers, TAWS uses numerical weather forecasts, real-world target models, and sensor characteristics to produce quantitative predictions of lock-on range. This is accomplished using physics-based models that predict weather effects on target contrast through thermal modeling of targets, backgrounds, and atmospheric transmission. TAWS supports strike mission planning by producing simulations of lock-on range versus the time of day or azimuthal angle of attack. The capability is currently being implemented in a mission planning initiative for generating air tasking order. The system enables weather impacts to be considered in weapon selection and time of attack planning. The result will be fewer weather aborts, improved effectiveness, reduced exposure to risk, and cost savings.

The program in atmospheric measurements and modeling for the airborne laser provides expanded understanding of atmospheric effects applicable to emerging HEL systems. High-altitude clouds (cirrus) and optical turbulence fundamentally limit ABL effectiveness and range. Early in the program, it was recognized that the variability of turbulence was producing variability in ABL performance. As a result, a parallel AFRL S&T program (with limited core S&T funding augmented by ABL funds) was initiated to examine atmospheric measurements and modeling. This effort has evolved into the Atmospheric Decision Aid (ADA) program.

From 1997 to 2000, the emphasis of the ADA program was on collecting theater turbulence data to validate ABL's design specification, which is based on data from core AFRL S&T work in the 1980s. Currently, the emphasis has shifted to modeling and forecasting. By merging the operational Air Force numerical weather model with an optical turbulence model, a three-dimensional forecast of turbulence,

including its temporal variability, is being developed. High-altitude clouds are another focus of the ADA model. More specifically, improved models of cirrus forecasting and the resulting laser transmission are required for ABL performance. AFRL is developing the models and software for delivery to the ABL's ADA integrator contractor to be implemented into a system that can be fielded.

The initial goal of the ADA is to support the ABL test phase by optimizing orbit placement; this modest goal can be achieved using existing models. The more ambitious goal, however, is to quantitatively forecast performance in terms of maximum effective range or required dwell time. To achieve this goal requires a longer-term core lab S&T program to improve the current state of the art of turbulence and cirrus modeling.

Emerging HEL systems for air-to-ground applications suffer far more performance variability than does the ABL in its missile defense mission. This is a result of the atmospheric boundary layer and the degree to which weather and diurnal cycle affect performance. In addition to turbulence, the full range of cloud fields and aerosols will need to be modeled in air-to-ground applications.

The first-order effect of clouds on electro-optics (EO) and HEL systems has to do with the line of sight to the target. This effect is quantified through the cloud-free line-of-sight (CFLOS) statistic. Although CFLOS is a basic concept, it is somewhat unique to military problems. Additionally, clouds vary significantly with season and location, and therefore a firm understanding of the climatology of CFLOS over militarily significant areas is needed. As future HEL systems are specified and designed, realistic physical models of clouds and CFLOS are required—particularly in light of the increased importance of virtual engineering, simulation, and testing. And as these systems are deployed, the ability to forecast clouds in terms of CFLOS probabilities will become essential. These capabilities will require an improvement in the current ability to forecast clouds, including improvements in satellite cloud sensing, using numerical weather models.

Although employment modes for some HEL systems may eliminate or minimize atmospheric effects (such as the space-based laser or reduced distances to the target for the ABL), it is desirable that these systems be effective in much more broadly defined scenarios, especially when adjunct missions are considered. Atmospheric modeling and decision aids will significantly enhance HEL systems and expand their operational capabilities, much like the demonstrated contribution of atmospheric decision aids to the effectiveness of comparatively simpler systems such as IR seekers. Like the Atmospheric Decision Aid (ADA) program for the ABL, expanded capabilities will need to be tailored to the operational scenarios and lethality mechanisms of new HEL systems. Advancements in atmospheric modeling and decision aids will require an expanded, long-term S&T program, as the need for atmospheric models is military specific and is not being addressed by the civilian research community. Further, a tri-service S&T program needs to coordinate the expertise resident across the service laboratories, so that it is effectively focused on this difficult problem.

8.6 Current Initiatives

The Department of Defense currently has three major initiatives underway involving high-energy lasers (HELs):

1. The airborne laser (ABL)
2. The space-based laser (SBL)
3. The Tactical High-Energy Laser (THEL)

The ABL is the furthest along in its development, having entered the program development and risk reduction phase in 1996. An initial operational capability for this boost-phase ballistic missile defense (BMD) system could be achieved in about 2010, following an aggressive testing schedule in the next few years. The Space-Based Laser project—a system also designed to destroy targets in the boost phase—is still in the very early development stage, with no decision at this time to pursue an operational system. Should a decision be made, initial operational capability is at least two decades away. The High-Energy Laser System-Tactical Army (HELSTAR), a US-only version of THEL, is under consideration to be funded as a new program. This system would provide HEL capability for the Army's Enhanced Area Air Defense System and other applications.

Each of these systems has the potential to contribute to multiple missions and provide a significant technological advantage to the warfighter. Technologies for high-energy lasers have matured to the point that a family of applications is feasible in the next few decades. A common thread in these initiatives, however, is the need for more robust science and technology investment to realize cost-effective operational capabilities.

Key S&T (scientific and technical) issues, which have an impact on all initiatives in this and following chapters, include pointing and tracking accuracy, beam control, and beam propagation in a battlefield environment or during poor weather conditions. In the case of laser weapons, lethality effects against a variety of targets must also be clearly understood. More specifically, these concerns are [31]:

- *Pointing and tracking accuracy*: is the ability to point the laser beam to the desired aimpoint and to maintain that aimpoint on the target.
- *Beam control*: refers to forming and shaping the beam. Depending on the nature of the specific laser, beam control can include initial processing of the beam to shape it and eliminate unwanted off-axis energy or can include wave-front shaping and/or phase control.
- *Beam propagation*: describes the effects on the beam after it leaves the HEL output aperture and travels through the battlefield environment to the target. Optical stability of the platform and beam interactions with the atmosphere, both molecules and aerosol particles, primarily determine the laser beam quality at the target. Beam quality is a measure of how effective the HEL is in putting its light into a desired spot size on the target.
- *Lethality*: defines the total energy and/or fluence level required to defeat specific targets. The laser energy must couple efficiently to the target, and it must exceed

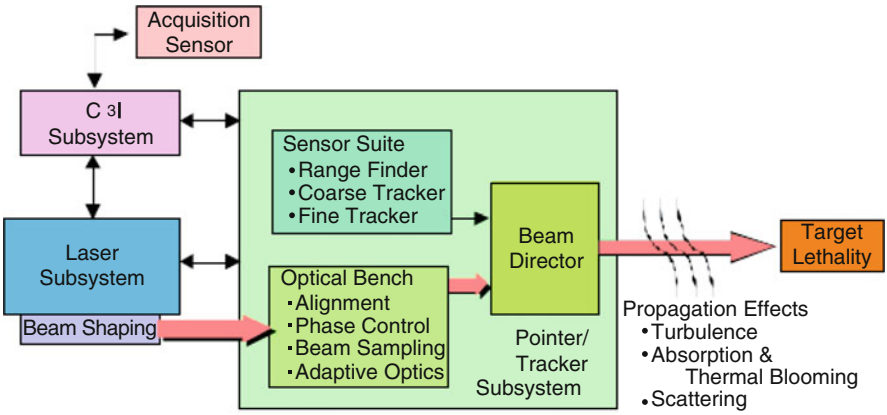


Fig. 8.24 Common high-energy laser technology [31]

a failure threshold that is both rate dependent and target specific. Laser output power and beam quality are two key factors for determining whether an HEL has sufficient fluence to negate a specific target, as Fig. 8.24 illustrates.

References

1. Tatarsky VI (1967) Wave propagation in the turbulent atmosphere. Nauka Publishers, Moscow (in Russian)
2. Gurvich S, Kon AI, Mironov VL, Khmelevtsov SS (1976) Laser radiation in the turbulent atmosphere. Nauka Publishers, Moscow (in Russian)
3. Rytov SM, Kravtsov YA, Tatarsky VI (1978) Introduction to the statistical physics, part 2. Nauka Publishers, Moscow (in Russian)
4. Semyonov A, Arsenyanf TI (1978) Fluctuations of electromagnetic waves on paths close to earth. Nauka Publishers, Moscow (in Russian)
5. Gochelashvi KS, Shishov VI (1981) Waves in randomly-inhomogeneous media, vol 1. Viniti Publishers, Moscow (in Russian)
6. Strohbeh JW (ed) (1978) Laser beam propagation in the atmosphere, Topics in applied physics. Springer, Berlin
7. Zuyev VI (1981) Propagation of laser radiation in the atmosphere. Radio i svyaz, Moscow (in Russian)
8. Ishimaruw A (1978) Wave propagation, and scattering in random media. Academic, New York
9. Akhmanov SA, Dyakov YE, Chirkin AS (1981) Introduction to the statistical radiophysics and optics. Nauka Publishers, Moscow (in Russian)
10. Mironov VL (1981) Propagation of a laser beam in the turbulent atmosphere. Nauka Publishers, Novosibirsk (in Russian)
11. Bakut PA, Ustinovn ND, Troitsky N, Sviridovz KN (1977) A rubezhnaya. Radioelectronika 1:3–29, 3, pp. 55-86 (in Russian)
12. Hardy JW (1978) Active optics: a new technology of the control of light. Proc IEEE 66 (6):651697
13. Lukin VP (1986) Atmospheric adaptive optics. Nauka Publishers, Novosibirsk (in Russian)

14. Raizer YP (ed) (1968) The effect of laser radiation. Collection of articles. Translated from English. Mir Publishers, Moscow (in Russian)
15. Akhmanovd SA, Krindacha P, Migulin V et al (1968) IEEE J Quantum Electron QE-4 (10):568–575
16. Gebhardt FG (1976) Appl Opt 15(6):1479–1493
17. Smith DC (1977) High power laser propagation, thermal blooming. Proc IEEE 65 (12):1679–1714
18. Gordin MP, Sokolov AV, Stielkov GM (1980) Ito gi nauki i tekhniki. Radiotekhnika 20:206–289, Viniti Publishers, Moscow
19. Zuyev VY, Kopytin YD, Kuzikovsky AV (1980) Nonlinear optic effects in aerosols. Nauka Publishers, Novosibirsk (in Russian)
20. Volkovitsky OA, Sedunov S, Semyonov LP (1982) Propagation of intensive laser radiation in clouds. Gidrometeoizdat Publishers, Leningrad (in Russian)
21. Goryachevf LV, Grigoryevv V, Kalinovsky V et al (1977) Quant Electron 4(4):907–909 (in Russian)
22. Averbakha VS, Betin A, Gapnov VA et al (1978) Radiojzika 8:1077–1106
23. Akhtyrchenko YV, Belyayev YB, Vysotsky YP et al (1983) Izv vuzov Fizika 26(2):5–13, Sov Phys J 26:105 (1983)
24. Glicler SL (1971) Appl Opt 10(3):644650
25. Sukhorukov P, Shljmilov EN (1973) Zh E T F 43(5):1029–1041
26. Loskutov VS, Strelkov GM (1984) Clearing of the polydispersed aerosol by the laser radiation. Proceedings of the all-union conference on propagation of the laser radiation in disperse media. Obninsk, Part 2, pp 162–163 (in Russian)
27. Brown RT, Smith DC (1975) J Appl Phys 46:402
28. Spangle P, Pefiano JR, Hafizi B (2007) Optimum wavelength and power for efficient laser propagation in various atmospheric environments. NRL/MR/6790-05-8907
29. Berger PJ, Ulrich PB, Ulrich JT, Gebhardt FG (1977) Appl Opt 16:345
30. Nielsen PE (1994) Effects of directed energy weapons. Philip E. Nielsen
31. Defense Science Board Task Force on High Energy Laser, Weapon System Applications, June 2001, Office of the Under Secretary of Defense For Acquisition, Technology, and Logistics. Washington, DC, 20301-3140
32. Mitchner M, Kruger CH Jr (1973) Partially ionized gases. Wiley Interscience, New York
33. Driscoll WG (ed) (1978) Handbook of optics. McGraw-Hill, New York, The upper portion of Figure 8.3 is based on a figure on p.115 of R. D. Hudson, Jr., Infrared Systems Engineering (New York: John Wiley and Sons, 1969). The lower, expanded portion of the figure is based on Figure 2 in Frederic G. Gebhardt, "High Power Laser Propagation, "Applied Optics 15, 1484 (1976). Gebhardt's paper is a good summary of many of the phenomena discussed in this chapter, at a somewhat higher level of technical detail
34. Reif F (1965) Fundamentals of statistical and thermal physics. McGraw-Hill, New York, Section 6.3
35. Zuev VE (1982) Laser beams in the atmosphere. Consultant's Bureau, New York
36. Born M, Wolf E (1975) Principles of optics, 5th edn. Pergamon Press, Oxford
37. Axaryan GA (1963) Zh E T F 45(9):810812 (in Russian). Sov Phys J ETP 18:555 (1964)
38. Burn VI, Chaporov DP (1975) Transparency dynamics of the solid aerosol in the interaction with the pulse radiation. Third all-union symposium on laser radiation propagation in the atmosphere. Abs. of Papers. Tom&: Institute of the Atmospheric Optics, pp 119–122 (in Russian).
39. Kolov VV, Chaporov DP (1983) Nonlinear distortion of the laser radiation in Haze. In: Zwev VY (ed) Problems of the atmospheric optics. Nauka Publishers, Novosibirsk, pp 3–12, in Russian
40. Vorobev VV (1991) Thermal blooming of laser beams in the atmosphere. The institute of Atmospheric Physics, U.S.S.R. Academy of Science. Prog Qmt Electr 15:I-152 (Pergamon Press plc)

41. Gebhardt FG (1976) High power laser propagation. Appl Opt 15:1484
42. Ulrich PB (1975) Numerical methods in high power laser propagation. AGARD conference proceedings No. 183, Optical Propagation in the Atmosphere, Paper No. 31, 27–31 October 1975
43. Bloembergen N, Patel CKN, Avizonis P, Clem RG, Hertzberg A, Johnson TH, Marshall T, Miller RB, Morrow WE, Salpeter EE, Sessler AM, Sullivan JD, Wyant JC, Yariv A, Zare RN, Glass AJ, Hebel LC, Pake GE, May MM, Panofsky WK, Schawlow AL, Townes CH, York H (1978) Report to The American Physical Society of the study group on science and technology of directed energy weapons. Rev Mod Phys 59(3 Part II), July 1978, Section 5.4.8.
44. Pearson JE (1976) Atmospheric turbulence compensation using coherent optical adaptive techniques. Appl Opt 15:622
45. Gebhardt FG (1976) High power laser propagation. Appl Opt 15(6):1479–1493

Appendix A
Short Course in Taylor Series

The Taylor series is mainly used for approximating functions when one can identify a small parameter. Expansion techniques are useful for many applications in physics, sometimes in unexpected ways.

A.1 Taylor Series Expansions and Approximations

In mathematics, the **Taylor series** is a representation of a function as an infinite sum of terms calculated from the values of its derivatives at a single point. It is named after the English mathematician Brook Taylor. If the series is centered at zero, the series is also called a **Maclaurin series**, named after the Scottish mathematician Colin Maclaurin. It is common practice to use a finite number of terms of the series to approximate a function. The Taylor series may be regarded as the limit of the Taylor polynomials.

A.2 Definition

A Taylor series is a *series expansion* of a *function* about a point. A one-dimensional Taylor series is an expansion of a real function $f(x)$ about a point $x = a$ is given by;

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^3(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots \tag{A.1}$$

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

If $a = 0$, the expansion is known as a *Maclaurin Series*. Equation A.1 can be written in the more compact sigma notation as follows:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{A.2}$$

where $n!$ is mathematical notation for factorial n and $f^{(n)}(a)$ denotes the n th derivation of function f evaluated at the point a. Not that the zeroth derivation of f is defined to be itself and both $(x-a)^0$ and $0!$ by their mathematical definitions are set equal to 1.

Taylor series of some common functions expanding around point $x = a$ include;

$$\frac{1}{1-x} = \frac{1}{1-a} + \frac{x-a}{(1-a)^2} + \frac{(x-a)^2}{(1-a)^2} + \frac{(x-a)^3}{(1-a)^3} + \cdots \tag{A.3a}$$

$$e^x = e^a\left\{1 + (x-a) + \frac{1}{2}(x-a)^2 + \frac{1}{6}(x-a)^3 + \cdots\right\} \tag{A.3b}$$

$$1\mathrm{n}x = 1\mathrm{n}a + \frac{x-a}{a} - \frac{(x-a)^2}{2a^2} + \frac{(x-a)^3}{3a^3} + \cdots \tag{A.3c}$$

$$\cos x = \cos a - \sin a(x-a) - \frac{1}{2}\cos a(x-a)^2 + \frac{1}{6}\sin a(x-a)^3 + \cdots \tag{A.3d}$$

$$\sin x = \sin a - \cos a(x-a) - \frac{1}{2}\sin a(x-a)^2 + \frac{1}{6}\cos a(x-a)^3 + \cdots \tag{A.3e}$$

$$\tan x = \tan a - \sec^2 a(x-a) - \sec^2 a \tan a(x-a)^2 + \sec^2 a\left(\sec 2a - \frac{2}{3}\right)(x-a)^3 + \cdots \tag{A.3f}$$

Derivation of any f these function can be found in any calculus books.

Taylor series can also be defined for functions of a complex variable. By the *Cauchy integral formula* and is written in form of;

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_C \frac{f(z')dz'}{z'-z} \\ &= \frac{1}{2\pi i}\int_C \frac{f(z')dz'}{(z'-z_0)-(z-z_0)} \\ &= \frac{1}{2\pi i}\int_C \frac{f(z')dz'}{(z'-z_0)\left(1-\dfrac{z-z_0}{z'-z_0}\right)} \end{aligned} \tag{A.4}$$

In the interior of C,

$$\frac{|z-z_0|}{|z'-z_0|} < 1$$

So, using

$$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n$$

It follows that;

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_C \sum_{n=0}^{\infty} \frac{(z-z_0)^n f(z')dz'}{(z'-z_0)^{n+1}} \\ &= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \int_C \frac{f(z')dz'}{(z'-z_0)^{n+1}} \end{aligned} \tag{A.5}$$

Using the Cauchy integral formula for derivatives,

$$f(z) = \sum_{n=0}^{\infty}(z-z_0)^n \frac{f^{(n)}(z_0)}{n!} \tag{A.6}$$

A.3 Maclaurin Series Expansions and Approximations

In the particular case where $a = 0$, the series is also called a **Maclaurin series**:

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^2 + \cdots \tag{A.7}$$

A.4 Derivation

The Maclaurin/Taylor series can be derived in the following manner.

If a function is analytic, it may be defined by a power series:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{A.8}$$

Evaluating at $x = 0$, we have:

$$f(0) = a_0 \tag{A.9}$$

Differentiating the function,

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \cdots \tag{A.10}$$

Evaluating at $x = 0$,

$$f'(0) = a_1$$

Differentiating the function again,

$$f''(x) = 2a_2 + 6a_3 x + 12a_4 x^2 + \cdots$$

Evaluating at $x = 0$,

$$\frac{f''(x)}{2!} = a_2$$

Generalizing,

$$a_n \frac{f^n(0)}{n!}$$

where $f^n(0)$ is the nth derivative of $f(0)$.

Substituting the respective values of a_n in the power expansion,

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^2 + \cdots \tag{A.11}$$

which is a particular case of the Taylor series (also known as Maclaurin series).

Generalizing further, by writing f in a more general form, allowing for a shift of a, we have

$$f(x) = b_0 + b_1(x-a) + b_2(x-a)^2 + b_3(x-a)^3 + \cdots \tag{A.12}$$

Generalizing similar as above, for evaluating at x = a, we get

$$b_n = \frac{f^n(a)}{n!} \tag{A.13}$$

Substituting, we get

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots \tag{A.14}$$

which is the Taylor series.

Examples The Maclaurin series for any polynomial is the polynomial itself.

The Maclaurin series for $(1-x)^{-1}$ is the geometric series

$$1 + x + x^2 + x^3 + \cdots \tag{A.15}$$

so the Taylor series for x^{-1} at $a = 1$ is

$$1 - (x-1) + (x-1)^2 - (x-1)^2 + \cdots \tag{A.16}$$

By integrating the above Maclaurin series we find the Maclaurin series for $1n(1-x)$, where ln denotes the natural logarithm:

$$x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots \tag{A.17}$$

and the corresponding Taylor series for $1n\,x$ at $a = 1$ is

$$(x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \cdots \tag{A.18}$$

The Taylor series for the exponential function e^x at $a = 0$ is

$$1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \cdots \tag{A.19}$$

The above expansion holds because the derivative of e^x with respect to x is also e^x and e^0 equals 1. This leaves the terms $(x-0)^n$ in the numerator and $n!$ in the denominator for each term in the infinite sum.

A.5 Convergence

The sine function is closely approximated by its Taylor polynomial of degree 7 for a full period centered at the origin.

The Taylor polynomials for $\log(1+x)$ only provide accurate approximations in the range $-1 < x \leq 1$. Note that, for $x > 1$, the Taylor polynomials of higher degree are ***worse*** approximations.

In general, Taylor series need not be convergent. More precisely, the set of functions with a convergent Taylor series is a meager set in the Frechet space of smooth functions. In spite of this, for many functions that arise in practice, the Taylor series does converge.

The limit of a convergent Taylor series of a function f need not in general be equal to the function value $f(x)$, but in practice often it is. For example, the function

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if} \quad x \neq 0 \\ 0 & \text{if} \quad x = 0 \end{cases} \tag{A.20}$$

is infinitely differentiable at x $= 0$, and has all derivatives zero there. Consequently, the Taylor series of $f(x)$ is zero. However, $f(x)$ is not equal to the zero function, and so it is not equal to its Taylor series.

If $f(x)$ is equal to its Taylor series in a neighborhood of a, it is said to be analytic in this neighborhood. If $f(x)$ is equal to its Taylor series everywhere it is called entire. The exponential function e^x and the trigonometric functions sine and cosine are examples of entire functions. Examples of functions that are not entire include the logarithm, the trigonometric function tangent, and its inverse arctan. For these functions the Taylor series do not converge if x is far from a.

Taylor series can be used to calculate the value of an entire function in every point, if the value of the function, and of all of its derivatives, are known at a single point. Uses of the Taylor series for entire functions include:

1. The partial sums (the Taylor Polynomials) of the series can be used as approximations of the entire function. These approximations are good if sufficiently many terms are included.
2. The series representation simplifies many mathematical proofs.

Pictured on the right is an accurate approximation of $\sin(x)$ around the point $a = 0$. The pink curve is a polynomial of degree seven:

$$\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} \tag{A.21}$$

The error in this approximation is no more than $|x|^9/9!$. In particular, for $-1 < x < 1$, the error is less than 0.000003.

In contrast, also shown is a picture of the natural logarithm function $\log(1+x)$ and some of its Taylor polynomials around $a = 0$. These approximations converge to the function only in the region $-1 < x \leq 1$; outside of this region the higher-degree Taylor polynomials are ***worse*** approximations for the function. This is similar to Runge's phenomenon.

The **error** incurred in approximating a function by its nth-degree Taylor polynomial, is called the **remainder** or *residual* and is denoted by the function $R_n(x)$. Taylor's theorem can be used to obtain a bound on the size of the remainder.

Appendix B
Short Course in Vector Analysis

In this study of Laser Physics and related electricity and magnetism, a great saving in complexity of notation may be accomplished by using the notation of vector analysis. The purpose of this section is to give a brief but self-contained exposition of basic vector analysis to bring to the forefront of physical ideas involved in equations and provide knowledge of the field that is required for treatment of electromagnetic wave and Maxwell's equations.

B.1 Definitions

In study a basic and elementary physics, several kinds of quantities we are faced with and need to understand in particular we need to distinguish between Vectors and Scalars. For our purpose we define each of them separately as follows.

1. Scalar
   A Scalar is a quantity that is completely characterized by the magnitude.
   Examples of scalars are numerous such as mass, time, volume, etc. A simple extension of the idea of scalar is a *Scalar Field*—a function of position that is completely specified by its magnitude at all points is space
2. Vector
   A vector is defined as follow.
   A *Vector* is a quantity that is completely characterized by its magnitude and direction.
   Examples of vectors that we can cite is position from a fixed origin, velocity, acceleration, force, etc. The generalization to a *Vector Field* gives a function of position that is completely specified by its magnitude and direction at all points in space.

More complicated kinds of quantities, such as tensor, may be also defined. Scalars and Vectors will, however, largely suffice for our purpose.

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

B.2 Vector Algebra

Vector algebra is very similar to scalar algebra that readers are familiar with it. In order to proceed with this development it is convenient to have a representation of vectors, for which purpose we introduce a Three-Dimensional Cartesian Coordinate system. This Three-Dimensional system will be denoted by the three variables x, y, z or, when it is more convenient, x_1, x_2, x_3. With respect to this coordinate system, a vector is specified by it $x-$, $y-$, and z− components. Thus a vector V is specified by its components V_x, V_y, V_z, where $V_x = |V|\cos\alpha_1$, $V_y = |V|\cos\alpha_2$ and $V_z = |V|\cos\alpha_3$, the α's being the angles between V and appropriate coordinate axes. The scalar $|V| = \sqrt{V_x^2 + V_y^2 + V_z^2}$ is the *magnitude* of the vector V, or its *length*. In the case of Vector Fields, each of the components is to be regarded as a function of X, *Y*, and *Z*. It should be emphasized at this point that we introduce a representation of the vectors with respect to a Cartesian coordinate system only for simplicity and ease of understanding; all of the definitions and operations are, in fact, independent of any special choice of coordinates.

The sum of two vectors is defined as the vector whose components are the sums of the corresponding components of the original vectors. Thus if $-\mathsf{C}$ is the sum of $-\mathsf{A}$ and $-\mathsf{B}$, we write

$$\mathsf{C} = \mathsf{A} + \mathsf{B} \tag{B.1}$$

and

$$C_x = A_x + B_x \qquad C_y = A_y + B_y \qquad C_z = A_z + B_z \tag{B.2}$$

This definition of the vector sum is completely equivalent to the familiar parallelogram rule for vector addition.

Vector subtraction is defined in terms of the negative of a vector, which is the vector whose components are the negative of the corresponding components of the original vector. Thus if $\mathbf{A}$ is a vector, $-\mathbf{A}$ is defined by

$$(-\mathbf{A})_x = -A_x \quad (-\mathbf{A})_y = -A_y \quad (-\mathbf{A})_z = -A_z \tag{B.3}$$

The operation of subtraction is then defined as the additional of the negative and is written

$$\mathbf{A}-\mathbf{B} = \mathbf{A}+(-\mathbf{B}) \tag{B.4}$$

Since the addition of real numbers is associative and commutative, it follows that vector addition (and subtraction) is also associative and commutative. In vector notation this appears as

$$\begin{aligned}\mathbf{A} + (\mathbf{B}+\mathbf{C}) &= (\mathbf{A}+\mathbf{B})+\mathbf{C} \\ &= (\mathbf{A}+\mathbf{C})+\mathbf{B} = \mathbf{A}+\mathbf{B}+\mathbf{C}\end{aligned} \tag{B.5}$$

In other words, the parentheses are not needed, as indicated by the last form.

Proceeding now to the process of multiplication, the simplest step is product of a scalar with a vector. The result of such operation is a vector and each component of which is the scalar times the corresponding component of the original vector. If c is a scalar and $\mathbf{A}$ is a vector, the product c$\mathbf{A}$ is a vector $\mathbf{B} = c\boldsymbol{A}$, defined by

$$B_{\mathrm{x}} = cA_x \quad B_{\mathrm{y}} = cA_y \quad B_z = cA_z \tag{B.6}$$

It is clear that if $\mathbf{A}$ is a *vector field* and $\boldsymbol{c}$ is a *scalar field*, then $\mathbf{B}$ is a new vector field that is *not* necessary a constant multiple of the original field.

If two vectors are to be multiplied, there are then two possibilities of occurrence and that is known as vector and scalar products of such multiplication. Considering first the scalar product, we note that this name derives from the scalar nature of the product, although the alternative names, *inner product* and *dot product*, are sometimes used as well. The definition of the scalar product, written $\mathbf{A} \cdot \mathbf{B}$ is

$$\mathbf{A} \cdot \mathbf{B} = A_xB_x + A_yB_y + A_zB_z \tag{B.7}$$

This definition is equivalent to another, and perhaps more familiar, definition—that is, as the product of the magnitudes of the original vectors times the cosine of the angles between these vectors. If $\mathbf{A}$ and $\mathbf{B}$ are perpendicular to each other, then;

$$\mathbf{A} \cdot \mathbf{B} = 0$$

The scalar product is commutative. The length of $\mathbf{A}$ is

$$|\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}}$$

The vector product of two vectors is a vector, which accounts for the name. Alternative names are *outer product* and *cross product*. The vector product is written as $\mathbf{A}\times\mathbf{B}$. If $\mathbf{C}$ is the vector product of $\mathbf{A}$ and $\mathbf{B}$, then $\mathbf{C} = \mathbf{A}\times\mathbf{B}$ or

$$C_x = A_yB_z - A_zB_y \quad C_y = A_zB_x - A_xB_z \quad C_z = A_xB_y - A_yB_x \tag{B.8}$$

Note: that above results can be obtained easily using a determinant form of the cross product of $\mathbf{C} = \mathbf{A}\times\mathbf{B}$ as follows;

$$\begin{vmatrix} c_x \\ c_y \\ c_z \end{vmatrix} \begin{vmatrix} \hat{i} \\ \hat{j} \\ \hat{k} \end{vmatrix} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ A_x & A_y & A_z \\ B_x & B_y & C_z \end{vmatrix} = \hat{i}\left(A_yB_z - A_zB_y\right) - \hat{j}\left(A_xB_z - A_zB_z\right) + \hat{k}\left(A_xB_y - A_yB_x\right)$$

$$= \hat{i}\left(A_yB_z - A_zB_y\right) - \hat{j}\left(A_zB_x - A_xB_z\right) + \hat{K}\left(A_xB_y - A_yB_x\right)$$

and by equating each side of equality sign to its own appropriate components we get the result in Eq. (B.8).

It is important to note that the cross product depends on the order of the factor; interchanging the order introduces a minus sign:

$$\mathbf{B}\times\mathbf{A} = -\mathbf{A}\times\mathbf{B}$$

Consequently,

$$\mathbf{A}\times\mathbf{A} = 0$$

This definition is equivalent to the following: The vector product is the product of the magnitudes times the sin of the angle between the original vectors, with the direction given by a right-hand screw rules (i.e. Let **A** be rotated into **B** through the smallest possible angle. A right-hand screw rotated in this manner will advance in a direction perpendicular to both **A** and **B**; this direction is the direction of **A**×**B**).

The vector product may be easily remembered in terms of a determinant. If $\hat{\imath}, \hat{j}$ and $\hat{k}$ are unit vectors—that is, vectors of unit magnitude, in the x-, y-, and z-direction, respectively—Then

$$\mathbf{A}\times\mathbf{B} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ A_x & A_y & A_z \\ B_x & B_y & C_z \end{vmatrix} \tag{B.9}$$

If this determinant is evaluated by the usual rules, the result is precisely our definition of the cross product.

The preceding algebraic operations may be combined in many ways. Most of the results so obtained are obvious; however, there are two triple products of sufficient importance to merit explicit. The triple scalar product

$$D = \mathbf{A} \cdot \mathbf{B} \times \mathbf{C} = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix} = -\mathbf{B} \cdot \mathbf{A} \times \mathbf{C} \tag{B.10}$$

This product is unchanged by an exchange of dot and cross or by a cyclic permutation of the three vectors.

Note: that parentheses are not needed, since the cross product of a *scalar* and a *vector* is undefined.

The other interesting observation is about triple product of vectors such as $\mathbf{D} = \mathbf{A} \times (\mathbf{B} \times \mathbf{C})$. By a repeated application of the definition of the cross product, Eq. (B.8), we find

$$\mathbf{D} = \mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \tag{B.11}$$

which is frequently known as the *back cab rule*. It should be noted that in the cross product the parentheses are vital; without them the product is not well defined.

At this point we now define vector division and try to expand on its operation. This operation can be defined in two folds as follows:

3. Division of a vector by a scalar.
4. Division of a vector by another vector.

Division of a vector by a scalar can, of course, be defined as multiplication by reciprocal of the scalar.

Division of a vector by another vector will be possible if the two vectors are parallel. In other words, it is possible to write general solutions to vector equations and so accomplish something closely akin to division.

Consider the equation

$$c = \mathbf{A} \cdot \mathbf{X} \tag{B.12}$$

where **c** is a known scalar, **A** is a known vector, and **X** is an unknown vector. General solution that satisfies Eq. (B.12) is given as follows:

$$\mathbf{X} = \frac{c\mathbf{A}}{\mathbf{A} \cdot \mathbf{A}} + \mathbf{B} \tag{B.13}$$

Where **B** is an arbitrary vector that is perpendicular to **A** that is satisfying the relationship of $\mathbf{A} \cdot \mathbf{B} = \mathbf{0}$. The above steps is very nearly to divide c by **A**; more correctly, we have found the general form of the vector that satisfies Eq. (B.12). There is no unique solution, and this fact accounts for the vector **B**. In the same fashion, we may consider the vector equation

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} \tag{B.14}$$

where **A** and **C** are known vectors and **X** is an unknown vector. The general solution of this equation is

$$\mathbf{X} = \frac{\mathbf{C} \times \mathbf{B}}{\mathbf{A} \cdot \mathbf{A}} + \mathrm{k}\mathbf{A} \tag{B.15}$$

where **k** is an arbitrary scalar. Thus, **X** as defined by Eq. (B.15) is very nearly the quotient of **C** by **A**; the scalar ***K*** takes account of the non-uniqueness of process. If **X** is required to satisfy both Eqs. B.12 and B.14, then the result is unique if there exists such matter and is given by:

$$\mathbf{X} = \frac{\mathbf{C} \times \mathbf{B}}{\mathbf{A} \cdot \mathbf{A}} + \frac{c\mathbf{A}}{\mathbf{A} \cdot \mathbf{A}} \tag{B.16}$$

B.3 Gradient

Since the previous section was dealing with some extension of differentiation and integration we will now consider the vector calculus. The simples of all of these relationships are known as a particular vector field to the derivatives of a scalar field. It is convenient first to introduce the idea of the *directional derivative* of a function of several variables, which is just the rate of change of the function in specified direction. The direction derivative of a scalar function φ is usually denoted by $d\varphi/ds$; it must be understood that $d\mathbf{s}$ represents an infinitesimal displacement in the direction being considered, and that ds is the scalar magnitude of $d\mathbf{s}$. If $d\mathbf{s}$ has the components dx, dy, dz, then

$$\begin{aligned} \frac{d\varphi}{ds} &= \lim_{\Delta s \to 0} \frac{\varphi(x + \Delta x, y + \Delta y, z + \Delta z)}{\Delta s} \\ &= \frac{\partial \varphi}{\partial x}\frac{dx}{ds} + \frac{\partial \varphi}{\partial y}\frac{dy}{ds} + \frac{\partial \varphi}{\partial z}\frac{dz}{ds} \end{aligned}$$

In order to clarify the idea of a directional derivative, consider a scalar function of two variables. Thus, $\varphi(x, y)$ represents a two-dimensional scalar field. We may plot φ as a function of x and y as is depicted in Fig. B.1 for the function $\varphi(x, y) = x^2 + y^2$. The directional derivative at the point $x_0 y_0$ depends on the direction. If we choose the direction corresponding to $dy/dx = -x_0/y_0$, then we find

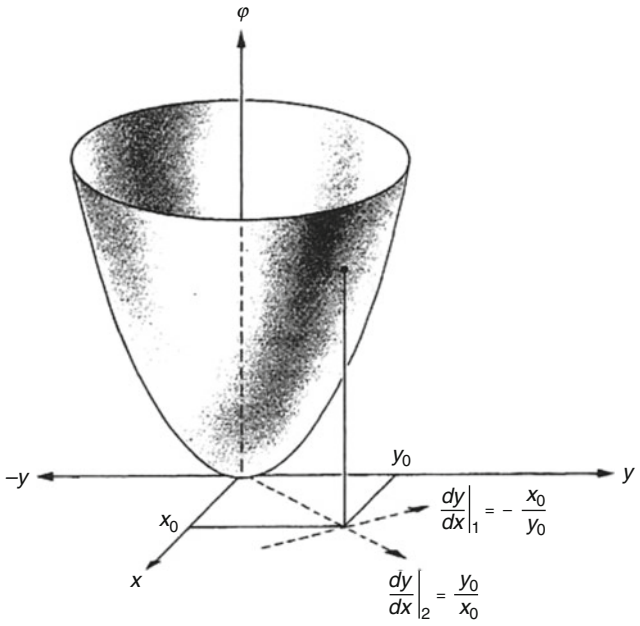


Fig. B.1 The function $\varphi(x, y) = x^2 + y^2$ plotted against x and y in a three-dimensional graph

$$\left.\frac{d\varphi}{ds}\right|_{x_0,y_0} = \frac{\partial\varphi}{\partial x}\frac{dx}{ds} + \frac{\partial\varphi}{\partial y}\frac{dy}{ds} = \left[\frac{\partial\varphi}{\partial x} + \frac{\partial\varphi}{\partial y}\frac{\frac{dy}{ds}}{\frac{dx}{ds}}\right]_{x_0,y_0} = \left[2x_0 - 2y_0\frac{x_0}{y_0}\right] = 0 \quad \text{(B.17a)}$$

Alternatively, if we choose $dy/dx = y_0/x_0$, we find

$$\begin{aligned}\left.\frac{d\varphi}{ds}\right|_{x_0,y_0} &= \left(2x_0 + 2\frac{y_0^2}{x_0}\right)\sqrt{\frac{x_0^2}{x_0^2 + y_0^2}} = 2\left(\frac{x_0^2 + y_0^2}{x_0}\right)\sqrt{\frac{x_0^2}{x_0^2 + y_0^2}} \\ &= 2\sqrt{\frac{(x_0^2 + y_0^2)x_0^2}{x_0^2(x_0^2 + y_0^2)}} = 2\sqrt{x_0^2 + y_0^2}\end{aligned} \quad \text{(B.17b)}$$

Since $ds = \sqrt{(dx)^2 + (dy)^2}$. As a third possibility, choose $dy/dx = \alpha$; then Eq. (B.17b) can be written as follows:

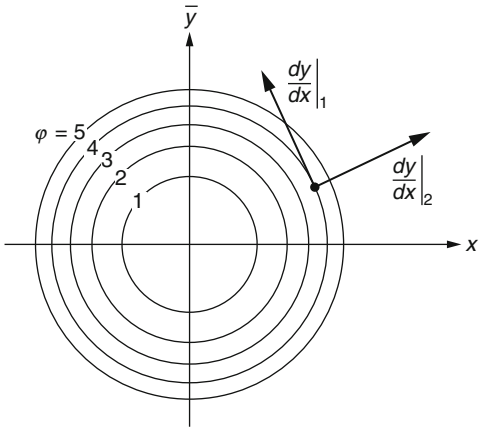


Fig. B.2 The function $\varphi(x, y)$ of Fig. B.1 expressed as a contour map in two dimensions

$$\left.\frac{d\varphi}{ds}\right|_{x_0,y_0} = \left(2x_0 + 2\frac{y_0^2}{x_0}\right)\sqrt{\frac{x_0^2}{x_0^2+y_0^2}} = \left(2x_0 + 2\frac{y_0}{x_0}y_0\right)\sqrt{\frac{x_0^2}{x_0^2\left(1+\frac{y_0^2}{x_0^2}\right)}}$$

$$= (2x_0 + 2\alpha y_0)\left(1+\alpha^2\right)^{1/2} \tag{B.17c}$$

In order to find value of α, we differentiate Eq. (B.17c) with respect to α and set it equal to zero for maximum or minimum of it. Performing these operations result in $\alpha = y_0/x_0$, which simply means that the direction of maximum rate of change of the function $\varphi = x^2 + y^2$ is the radial direction. If the direction is radially outward, then the maximum is the maximum rate of increase; if it is radially inward, it is a maximum rate of decrease or minimum rate of increase. In the direction specified by $dy/dx = -x_0/y_0$, the rate of change of $x^2 + y^2$ is zero. This direction is tangent to the circle $x^2 + y^2 = x_0^2 + y_0^2$. Clearly, on this curve, $\varphi = x^2 + y^2$ does not change. The direction in which $d\varphi/ds$ vanishes gives the direction of the curve $\varphi = constant$ through the point being considered. These lines, which are circles for the function $x^2 + y^2$, are completely analogous to the familiar contour lines or lines of constant altitude that appear on topographic maps. Figure B.2 shows the function $\varphi = x^2 + y^2$ re-plotted as a contour map.

The idea of contour lines may be generalized to a function of three variables, in which case the surfaces, $\varphi(x, y, z) = constant$, are called *level surface* or *equipotential surfaces*. The three-dimensional analog to Fig. B.2 is the only practical way of graphing a scalar field for a three-dimensional space. The gradient of a scalar function may now be defined as follows.

Note: The *gradient* of a scalar function φ is a vector whose magnitude is the maximum directional derivative at the point being considered and whose direction is the direction of the maximum directional derivative at the point.

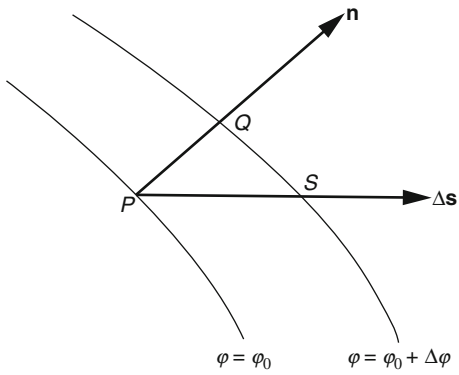


Fig. B.3 Parts of two level surfaces of the function $\varphi(x, y, z)$. $|\mathbf{grad}\varphi|$ at P equals the limit as $PQ \to 0$ of $\Delta\varphi/PQ$ and $d\boldsymbol{\varphi}/ds$ is the corresponding limit of $\Delta\varphi/\overline{PS}$

It is evident that the gradient has the direction normal to the level surface of φ through the point being considered. The most common symbols for the gradient are ∇ and **grad**. In terms of the gradient, the directional derivative is given by

$$\frac{d\varphi}{ds} = |\mathbf{grad}\ \varphi| \cos\theta \tag{B.18}$$

Where θ is the angle between the direction of $d\mathbf{s}$ and the direction of the gradient. This result is immediately evident from the geometry of Fig. B.3. If we write $d\mathbf{s}$ for the vector displacement of magnitude ds, then Eq. (B.18) can be written as;

$$\frac{d\varphi}{ds} = \mathbf{grad}\ \varphi \cdot \frac{d\mathbf{s}}{ds} \tag{B.19}$$

This equation enables us to find the explicit form of the gradient in any coordinate system in which we know the form of $d\mathbf{s}$. In rectangular coordinates, we know that $d\mathbf{s} = \mathbf{i}dx + \mathbf{j}dy + \mathbf{k}dz$. We also know that

$$d\varphi = \frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy + \frac{\partial\varphi}{\partial z}dz$$

From this equation and Eq. (B.19), it follows that

$$\frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy + \frac{\partial\varphi}{\partial z}dz = (\mathbf{grad}\ \varphi)_x dx + (\mathbf{grad}\ \varphi)_y dy + (\mathbf{grad}\ \varphi)_z dz$$

Equating coefficients of differentials of independent variables on both sides of the equation gives;

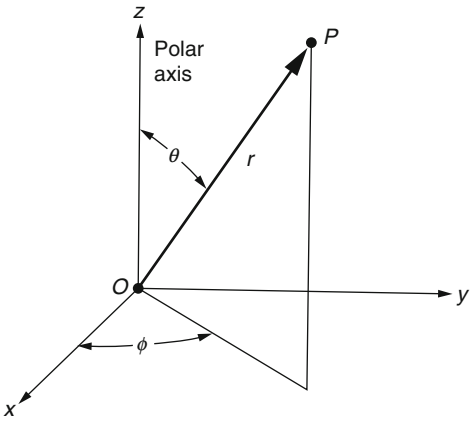


Fig. B.4 Definition of the polar coordinates r, θ and ϕ

$$\mathbf{grad}\ \varphi = \mathrm{i}\frac{\partial \varphi}{\partial x} + \mathrm{j}\frac{\partial \varphi}{\partial y} + \mathrm{k}\frac{\partial \varphi}{\partial z} \tag{B.20}$$

In rectangular coordinates. In a more complicated case, the procedure is the same. In spherical polar coordinates, with r, θ, ϕ as defined in Fig. B.4, we have

$$d\varphi = \frac{\partial \varphi}{\partial r}dr + \frac{\partial \varphi}{\partial \theta}d\theta + \frac{\partial \varphi}{\partial \phi}d\phi \tag{B.21}$$

and

$$d\mathbf{s} = \mathbf{a_r}dr + \mathbf{a_\theta} r d\theta + \mathbf{a_\phi} r \sin\theta d\phi \tag{B.22}$$

Where $\mathbf{a}_\mathrm{r}$, $\mathbf{a}_\theta$ and $\mathbf{a_\phi}$ are unit vectors in the r, θ, and ϕ directions, respectively. Applying Eq. (B.19) and equating coefficients of independent variables yields

$$\frac{d\varphi}{ds} = \mathbf{grad}\ \varphi \cdot \frac{d\mathbf{s}}{ds} = \mathbf{grad}\ \varphi \cdot \frac{\mathbf{a_r}dr + \mathbf{a_\theta} r d\theta + \mathbf{a}_\phi r \sin\theta d\phi}{ds}$$

$$\mathbf{grad}\ \varphi = \mathbf{a_r}\frac{\partial \varphi}{\partial r} + \mathbf{a_\theta}\frac{1}{r}\frac{\partial \varphi}{\partial \theta} + \mathbf{a_\varphi}\frac{1}{r\sin\theta}\frac{\partial \varphi}{\partial \varphi} \tag{B.23}$$

in spherical coordinates

B.4 Vector Integration

Although there are other aspects to differentiation involving vectors, it is convenient to discuss vector integration first. For our purposes, we may consider three kinds of integrals according to the nature of differential appearing in the integral as follows:

1. Line
2. Surface
3. Volume

The integrand may be either a vector field or a scalar field; however, certain combinations of integrands and differentials give rise to uninteresting integrals. Those of most interests are the scalar line integral of a vector, the scalar surface integral of a vector, and the volume integrals of both vectors and scalars.

If **F** is a vector field, a line integral of **F** is written

$$\int_{a(C)}^{b} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{l} \tag{B.24}$$

Where C is the curve along which the integration is performed, a and b the initial and final points on the curve C. Since F □ dl is a scalar, then it is clear that the line integral is a scalar. The definition of the line integral follows closely the Riemann definition of the definite integral. The segment of C between a and b is divided into a large number of small increments $\Delta \mathrm{l}_i$; for each increment an interior point is chosen and the value of **F** at that point found. The scalar product of each increment with the corresponding value of **F** is found and the sum of these computed. The line integral is then defined as the limit of this sum as the number of increments becomes infinite in such a way that each increment goes to zero. This definition may be compactly written as

$$\int_{a(C)}^{b} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{l} = \lim_{N \to \infty} \sum_{i=1}^{N} \mathbf{F}_i \cdot \Delta \mathbf{l}_i$$

It is important to note that the line integral usually depends not only on the endpoints a and b but also on the curve C along which the integration is to be done, since the magnitude and direction of **F**(**r**) and the direction of d**l** depend on C and its tangent, respectively. The line integral around a closed curve is of sufficient importance that a special notation is used for it, namely,

$$\oint_{C} \mathbf{F} \cdot d\mathbf{l} \tag{B.25}$$

The integral around a closed curve is usually not zero; the class of vectors for which the line integral around any closed curve is zero is of considerable importance. For this reason, one often encounters line integrals around undesignated closed paths. For example,

$$\oint \mathbf{F} \cdot d\mathbf{l} \tag{B.26}$$

This notation is useful only in those cases where the integral is independent of the contour C within rather wide limits. If any ambiguity is possible, it is wise to specify the contour. The basic approach to the evaluation of line integrals is to obtain a one-parameter description of the curve and then use this description to express the line integral as the sum of three ordinary one-dimensional integrals. In all but the simplest cases, this procedure is long and tedious. Fortunately, it is seldom necessary to evaluate the integrals in this fashion. As will be seen later, it is often possible to show that the line integral does not depend on the path between the endpoints. In the latter case, a simple path may be chosen to simplify the integration.

If $\mathbf{F}$ is again a vector, a surface integral of $\mathbf{F}$ is written

$$\int_S \mathbf{F} \cdot \mathbf{n}\, da \tag{B.27}$$

where S is the surface over which the integration is to be performed, da is an infinitesimal area on S and $\mathbf{n}$ is a unit normal to da. There is a twofold ambiguity in the choice of $\mathbf{n}$, which is resolved by taking $\mathbf{n}$ to be outward drawn normal if S is a closed surface. If S is not closed and is finite, then it has a boundary, and the sense of the normal is important only with respect to the arbitrary positive sense of traversing the boundary. The positive sense of the normal is the direction in which a right-hand screw would advance if rotate in the direction of the positive sense on the bounding curve, as illustrated in Fig. B.5. The surface integral of $\mathbf{F}$ over a closed surface S is sometimes denoted by

$$\oint_S \mathbf{F} \cdot \mathbf{n}\, da$$

Comments exactly parallel to those made for the line integral can be made for the surface integral. This surface integral is clearly a scalar; it usually depends on the surface S, and cases where it does not are particularly important. The definition of the surface integral is made in a way exactly comparable to that of the line integral. The detailed formulation is left for the problem section.

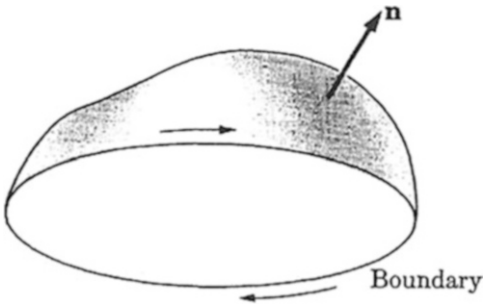


Fig. B.5 Relation of normal **n** to a surface and the direction of traversal of the boundary

If **F** is a vector and φ a scalar, then the two volume integrals in which we are interested are

$$J = \int_v \varphi \, dv \text{ and } \mathbf{K} = \int_v \mathbf{F} \, dv \tag{B.28}$$

Clearly J is a scalar and **K** is a vector. The definitions of these integrals reduce quickly to just the Riemann integral in three dimensions except that in **K** one must note that there is one integral for each component of **F**. These integrals are sufficiently familiar to require no further comment.

B.5 Divergence

Another important operator, which is essentially a derivative, is the divergence operator. The divergence of vector **F**, written div **F**, is defined as follows.

The *divergence* of a vector is the limit of its surface integral per unit volume as the volume enclosed by the surface goes to zero. This is:

$$\operatorname{div} \mathbf{F} = \lim_{V \to 0} \frac{1}{V} \oint_S \mathbf{F} \cdot \mathbf{n} \, da$$

The y a scalar point function (scalar field), and it is defined at the limit point of the surface of integration. This definition has several virtues: It is independent of any special choice of coordinate system, and it may be used to find the explicit form of the divergence operator in any particular coordinate system.

In rectangular coordinates the volume element $\Delta x \Delta y \Delta z$ provides a convenient basis for finding the explicit form of the divergence. If one corner of the rectangular parallelepiped is at the point x_0, y_0, z_0, then

$$F_x(x_0+\Delta x,y,z)=F_x(x_0,y,z)+\Delta x\frac{\partial F_X}{\partial x}\bigg|_{x_0,y,z}$$

$$F_y(x,y_0+\Delta y,z)=F_y(x,y_0,z)+\Delta y\frac{\partial F_Y}{\partial Y}\bigg|_{x,y_0,z} \tag{B.29}$$

$$F_z(x,y,z+\Delta z)=F_Z(x,y,z_0)+\Delta z\frac{\partial F_z}{\partial z}\bigg|_{x,y,z_0}$$

where higher-order terms in $\Delta x, \Delta y$, and Δz have been omitted. Since the area element $\Delta y\Delta z$ is perpendicular to the x -axis, $\Delta z\Delta x$ is perpendicular to the y -axis, and $\Delta x\Delta y$ is perpendicular to z -axis, the definition of the divergence becomes,

$$\begin{aligned}\operatorname{div}\mathbf{F}=\lim_{V\to 0}\frac{1}{\Delta x\Delta y\Delta z}\Bigg\{&\int F_x(x_0,y,z)\,dydz\\&+\Delta x\Delta y\Delta z\frac{\partial F_x}{\partial x}+\int F_y(x,y_0,z)\,dxdy\\&+\Delta x\Delta y\Delta z\frac{\partial F_y}{\partial y}+\int F_z(x,y,z_0)\,dxdy\\&+\Delta x\Delta y\Delta z\frac{\partial F_z}{\partial z}+\int F_x(x_0,y,z)\,dydz\\&-\int F_y(x,y_0,z)dxdz-\int F_z(x,y,z_0)\,dxdy\Bigg\}\end{aligned} \tag{B.30}$$

The minus signs associated with the last three terms account for the fact that the outward drawn normal is in the direction of the negative axes in these cases. The limit is easily taken, and the divergence in rectangular coordinates is found to be

$$\operatorname{div}\mathbf{F}=\frac{\partial F_x}{\partial x}+\frac{\partial F_y}{\partial y}+\frac{\partial F_z}{\partial z} \tag{B.31}$$

In spherical coordinates, the procedure is similar. The volume enclosed by the coordinate intervals $\Delta r, \Delta\theta, \Delta\phi$. Because the area enclosed by the coordinate intervals depends on the values of the coordinates (which is not the case with rectangular coordinates), it is best to write $\mathbf{F}\cdot\mathbf{n}\,\Delta a$ in its explicit form:

$$\begin{aligned}\mathbf{F}\cdot\mathbf{n}\Delta a=&F_r r^2\sin\theta\Delta\theta\Delta\phi\\&+F_\theta r\sin\theta\Delta\phi\Delta r+F_\phi r\Delta r\Delta\theta\end{aligned} \tag{B.32}$$

It is clear from this expression that $F_r r^2 \sin\theta$, rather than just F_r, must be expanded in Taylor series. Similarly, it is the coefficient of the products of coordinate intervals that must be expanded in the other terms. Making these expansions and using them to evaluate the surface integral in the definition of the divergence gives

$$\begin{aligned}\operatorname{div}\mathbf{F} = \lim_{V\to 0}\frac{1}{r^2\sin\theta\Delta r\Delta\theta\Delta\phi}\Bigg\{&\frac{\partial}{\partial r}\left(F_r r^2\sin\theta\right)\Delta r\Delta\theta\Delta\phi \\ &+\frac{\partial}{\partial\theta}\left(F_\theta r\sin\theta\right)\Delta\theta\Delta r\Delta\phi + \frac{\partial}{\partial\phi}\left(F_\phi r\right)\Delta\phi\Delta r\Delta\theta\Bigg\}\end{aligned} \tag{B.33}$$

Taking the limit, the explicit form of the divergence in spherical coordinates is found to be

$$\operatorname{div}\mathbf{F} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2F_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta F_\theta\right) + \frac{1}{r\sin\theta}\frac{\partial F_\phi}{\partial\phi} \tag{B.34}$$

This method of finding the explicit form of the divergence is applicable to any coordinate system, provided that the forms of the volume and surface elements or, alternatively, the elements of length are known.

The physical significance of the divergence is readily seen in terms of an example taken from fluid mechanics. If $\mathbf{V}$ is the velocity of a fluid, given as a function of position, and ρ is its density, then $\oint_S \rho\mathbf{V}\cdot\mathbf{n}\,da$ is clearly the net amount of fluid per unit time that leaves the volume enclosed by S. If the fluid is incompressible, the surface integral measures the total sources of fluid enclosed by the surface. The preceding definition of the divergence then indicates that it may be interpreted as the limit of the source strength per unit volume, or the source density of an incompressible fluid.

An extremely important theorem involving the divergence may be stated and proved.

Divergence theorem. The integral of the divergence of a vector over a volume V is equal to the surface integral of the normal component of the vector over the surface bounding V. That is:

$$\int_V \operatorname{div}\mathbf{F}\,dv = \oint_S \mathbf{F}\cdot\mathbf{n}\,da$$

Consider the volume to be subdivided into a large number of small cells. Let the ith cell have volume ΔV_i and be bounded by the surface S_i. It is clear that

$$\sum_i \oint_{S_i} \mathbf{F} \cdot \mathbf{n}\, da = \oint_S \mathbf{F} \cdot \mathbf{n}\, da \tag{B.35}$$

Where in each integral on the left, the normal is directed outward from the volume being considered. Since outward to one cell is inward to the appropriate adjacent cell, all contributions to the left side of Eq. (B.35) cancel except those arising from the surface of S, and Eq. (B.35) is essentially proved. The divergence theorem is now obtained by letting the number of cells go to infinity in such a way that the volume of each cell goes to zero.

$$\oint_{S_i} \mathbf{F} \cdot \mathbf{n}\, da = \lim_{\Delta V \to 0} \sum_i \left\{ \frac{1}{\Delta V_i} \oint_S \mathbf{F} \cdot \mathbf{n} da \right\} \Delta V_i \tag{B.36}$$

In the limit, the sum on i becomes an integral over V, and the ratio of the integral over S_i to ΔV_i becomes the divergence of $\mathbf{F}$. Thus,

$$\int_V \operatorname{div} \mathbf{F}\, dv = \oint_S \mathbf{F} \cdot \mathbf{n}\, da \tag{B.37}$$

which is the divergence theorem. We shall have frequent occasion to exploit this theorem, both in the development of the theoretical aspects of electricity and magnetism and for the very practical purpose of evaluating integrals.

B.6 Curl

The third interesting vector differential operator is the curl. The curl of a vector, written **curl F**, is defined as follows.

The *curl* of a vector is the limit of the ratio of the integral of its cross product with the outward drawn normal, over a closed surface, to the volume enclosed by the surface as the volume goes to zero. That is,

$$\mathbf{curl}\, \mathbf{F} = \lim_{V \to 0} \frac{1}{V} \oint_S \mathbf{n} \times \mathbf{F} \cdot da \tag{B.38}$$

The parallelism between this definition and the definition of the divergence is quite apparent; instead of the scalar product of the vector with the outward drawn normal, one has the vector product. Otherwise the definitions are the same. A different but equivalent definition is more useful. This alternative is as follows.

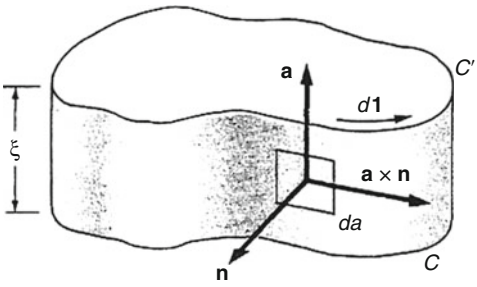


Fig. B.6 Volume swept out by displacing the plane curve C in the direction of its normal a

The component of **curl F** in the direction of the unit vector **a** is the limit of a line integral per unit area, as the enclosed area goes to zero, this area being perpendicular to **a**. That is,

$$\mathbf{a} \cdot \mathbf{curl}\,\mathbf{F} = \lim_{S \to 0} \oint_C \mathrm{F} \cdot d\mathbf{l} \tag{B.39}$$

where the curve C, which bounds the surface S, is in a plane normal to **a**.

It is easy to see the equivalence of the two definitions by considering a plane curve C and the volume swept out by this curve when it is displaced a distance ξ in the direction of the normal to its plane, as shown in Fig. B.6. If **a** is a normal to this plane, then taking the dot product of **a** with the first definition of the curl, Eq. (B.38), gives

$$\mathbf{a} \cdot \mathbf{curl}\,\mathbf{F} = \lim_{V \to 0} \frac{1}{V} \oint_S \mathbf{a} \cdot \mathbf{n} \times \mathbf{F} \cdot da$$

Since **a** is parallel to the normal for all of the bounding surface except the narrow strip bounded by **C** and $C^{'}$, only the integral over this surface need be considered. For this surface we note that $\mathbf{a} \times \mathbf{n}\, da$ is just $\xi d\mathbf{1}$, where $d\mathbf{1}$ is an infinitesimal displacement along C. Since, in addition $V = \xi S$, the limit of the volume integral is just

$$\mathbf{a} \cdot \mathbf{curl}\,\mathbf{F} = \lim_{V \to 0} \frac{1}{\xi S} \oint_S \xi \mathbf{F} \cdot d\mathbf{1}$$

which reduces to the second form of our definition upon canceling the ξ's. This equivalence can be shown without the use of the special volume used here; however, so doing scarifies much of the simplicity of the proof given above.

The form of the curl in various coordinate systems can be calculated in much the same way as was done with the divergence. In rectangular coordinates, the volume $\nabla x \Delta y \Delta z$ is convenient. For the $x-$ component of the curl only the faces perpendicular to the $y-$ and $z-$ axes contribute. Recalling that $\mathbf{j} \times \mathbf{k} = -\mathbf{k} \times \mathbf{j} = \mathbf{i}$, the non-vanishing contributions from the faces of the parallelepiped to the x-component of the curl give

$$(\mathbf{curl}\,\mathbf{F})_x = \lim_{V\to 0}\frac{1}{V}\{\left[-F_y(x,y,z+\Delta z)+F_y(x,y,z)\right]\Delta x\Delta y + [F_z(x,y+\Delta y,z)-F_z(x,y,z)]\Delta x\Delta y\} \tag{B.40}$$

Making a Taylor series expansion and taking the limit gives

$$(\mathbf{curl}\,\mathbf{F})_x = \frac{\partial F_x}{\partial y}-\frac{\partial F_y}{\partial z} \tag{B.41}$$

for the $x-$ component of the curl. The $y-$ and $z-$ components may be found in exactly the same way. They are

$$(\mathbf{curl}\,\mathbf{F})_y = \frac{\partial F_x}{\partial z}-\frac{\partial F_z}{\partial y} \quad \text{and} \quad (\mathbf{curl}\,\mathbf{F})_z = \frac{\partial F_y}{\partial x}-\frac{\partial F_x}{\partial y} \tag{B.42}$$

The form of the curl in rectangular coordinates can be easily remembered if it is noted that it is just the expansion of a three-by-three determinant namely,

$$\mathbf{curl}\,\mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} \tag{B.43}$$

Finding the form of the curl in other coordinate system is only slightly more complicated and is left to the problem section.

As with the divergence, we encounter an important and useful theorem involving the curl, known as Stokes's theorem.

Stokes's theorem: The line integral of a vector around a closed curve is equal to the integral of the normal component of its curl over any surface bounded by the curve. That is,

$$\oint_C \mathbf{F}\cdot d\mathbf{l} = \int_S \mathbf{curl}\,\mathbf{F}\cdot\mathbf{n}\,da \tag{B.44}$$

where C is a closed curve that bounds the surface S.

The proof of this theorem is quite analogous to the proof of the divergence theorem. The surface S is divided into a large number of the cells. The surface of the i th cell is called ΔS_i and the curve bounding it is C_i. Since each of these cells must be traversed in the same sense, it is clear that the sum of the line integrals around the

C_i's is just the line integral around the bounding curve; all of the other contributions cancel. Thus

$$\oint_C \mathbf{F} \cdot d\mathbf{l} = \sum_i \oint_{C_i} \mathbf{F} \cdot d\mathbf{l}$$

It remains only to take the limit as the number of cells becomes infinite in such a way that the area of each goes to zero. The results of this limiting process is

$$\oint_C \mathbf{F} \cdot d\mathbf{l} = \lim_{\Delta S_i \to 0} \sum_i \frac{1}{\Delta S_i} \oint_{C_i} \mathbf{F} \cdot d\mathbf{l}\, \Delta S_i = \int_S \mathbf{curl}\, \mathbf{F} \cdot \mathbf{n}\, da$$

which is Stokes's theorem. This theorem, like the divergence theorem, is useful both in the development of electromagnetic theory and the evaluation of integrals. It is perhaps worth noting that both the divergence theorem and Stokes's theorem are essentially partial integrations.

B.7 The Vector Differential Operator ∇

We now introduce an alternative notation for the three types of vector differentiation that have been discussed-namely, gradient, divergence, and curl. This notation uses the vector differential operator *del*, defined in Cartesian coordinates as

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} \tag{B.45}$$

Del is a differential operator in that it is used only in front of a function of (x, y, z), which it differentiates; it is a vector in that it obeys the laws of vector algebra*. In terms of del, Eqs. B.20, B.31, and B.44 are expressed as follows:

$$\mathbf{grad} = \nabla,$$

$$\nabla\varphi = \mathbf{i}\frac{\partial\varphi}{\partial x} + \mathbf{j}\frac{\partial\varphi}{\partial y} + \mathbf{k}\frac{\partial\varphi}{\partial z} \tag{B.20}$$

$$\mathrm{div} = \nabla\cdot,$$

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \tag{B.31}$$

$$\text{curl} = \nabla \text{x}\,,$$

$$\text{curl} = \nabla \times \text{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} \tag{B.44}$$

The operations expressed with del are themselves independent of any special choice of coordinate system. Any identities that can be proved using the Cartesian representation hold independently of the coordinate system. Del can be expressed in a non-Cartesian (curvilinear) orthonormal coordinate system in a form analogous to Eq. (B.46) with the appropriate distance elements, but it must be remembered in applying it that the unit vectors in such coordinate systems are themselves functions of position and have to be differentiated**. The important integral theorems, according to

***Note:** It is also a vector in terms of its transformation properties as shown

****Note:** Results for cylindrical and spherical coordinates are found in Equations B.19, B.45, and B.37, are

$$\int_{a(C)}^{b} \nabla\varphi \cdot d\mathbf{l} = \int_{a}^{b} d\varphi = \varphi\big|_a^b = \varphi_b - \varphi_a \tag{B.46}$$

$$\int_S \nabla \times \mathbf{F} \cdot \mathbf{n}\, da = \oint_S \mathbf{F} \cdot \mathbf{n}\, da \tag{B.45}$$

$$\int_V \nabla \cdot \mathbf{F}\, \mathrm{d}v = \oint_S \mathbf{F} \cdot \mathbf{n}\, da \tag{B.37}$$

These give the integral of a derivative of a function over a region of n dimensions, in terms of the values of the function itself on the $(n-1)$-dimensional boundary of the region, for $n = 1,2,3$. Because the del operator obeys the rules of vector algebra, it is convenient to use it in calculations involving vector analysis, and henceforth we shall express the gradient, divergence, and curl in terms of ∇. It should be noted that ∇ is a *linear* operator:

$$\nabla(a\varphi + b\psi) = a\nabla\varphi + b\nabla\psi$$

$$\nabla \cdot (a\mathbf{F} + b\mathbf{G}) = a\nabla \cdot \mathbf{F} + b\nabla \cdot \mathbf{G}$$

$$\nabla \times (a\mathbf{F} + b\mathbf{G}) = a\nabla \times \mathbf{F} + b\nabla \times \mathbf{G}$$

If a and b are constant scalars.

B.8 Further Developments

The gradient, divergence, and curl operations may be repeated on appropriate kinds of fields. For example, it makes sense to take the divergence of the gradient of scalar field. Indeed, this combined operation is sufficiently important that it has a special name, The *Laplacian*. It does not, however, make sense to take the curl of the divergence of a vector field because doing so involves taking the curl of a scalar, which is undefined. There is altogether five second-order operations that make sense, and two of these yield zero. All five, however, are very important in the study of electromagnetic fields.

The first of these is the *Laplacian Operator*, which is defined as the divergence of the gradient of a scalar field, and which is usually written ∇^2,

$$\nabla \cdot \nabla = \nabla^2$$

In rectangular coordinates,

$$\nabla^2 \varphi = \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z_2} \tag{B.47}$$

This operator is of great importance in electronics and heat transfer and laser effects on materials and is considered in Chap. 5.

The curl of the gradient of any scalar field is zero. This statement is most easily verified by writing it out in rectangular coordinate. If the scalar field is φ, then

$$\nabla \times (\nabla \varphi) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \dfrac{\partial \varphi}{\partial x} & \dfrac{\partial \varphi}{\partial x} & \dfrac{\partial \varphi}{\partial x} \end{vmatrix} = \mathbf{i}\left(\frac{\partial^2 \varphi}{\partial y \partial z} - \frac{\partial^2 \varphi}{\partial y \partial z}\right) + \ldots\ldots = 0 \tag{B.48}$$

which verifies the original statement. In operator notation,

$$\nabla \times \nabla = 0$$

The divergence of any curl is also zero. This result is verified directly in rectangular coordinates by writing

$$\nabla \times (\nabla \times \mathbf{F}) = \frac{\partial}{\partial x}\left(\frac{\partial F_x}{\partial y} - \frac{\partial F_y}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) + \ldots\ldots = 0 \tag{B.49}$$

Table B.1 Differential vector identification

$\nabla \times \nabla\varphi = \nabla^2\varphi$	Eq. (B.1.1.1)
$\nabla \cdot \nabla \times \mathbf{F} = 0$	Eq. (B.1.1.2)
$\nabla \times \nabla\varphi = 0$	Eq. (B.1.1.3)
$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2\mathbf{F}$	Eq. (B.1.1.4)
$\nabla(\varphi\psi) = (\nabla\varphi)\psi + \varphi\nabla\psi$	Eq. (B.1.1.5)
$\nabla(\mathbf{F}\cdot\mathbf{G}) = (\mathbf{F}\cdot\nabla)\mathbf{G} + \mathbf{F}\times(\nabla\times\mathbf{F}) + (\mathbf{G}\cdot\nabla)\mathbf{F}+\mathbf{G}\times(\nabla\times\mathbf{F})$	Eq. (B.1.1.6)
$\nabla\cdot(\varphi\mathbf{F}) = (\nabla\varphi)\cdot\mathbf{F} + \varphi\nabla\cdot\mathbf{F}$	Eq. (B.1.1.7)
$\nabla\cdot(\mathbf{F}\times\mathbf{G}) = (\nabla\times\mathbf{F})\cdot\mathbf{G} - (\nabla\times\mathbf{G})\cdot\mathbf{F}$	Eq. (B.1.1.8)
$\nabla\times(\varphi\mathbf{F}) = (\nabla\varphi)\times\mathbf{F} + \varphi\nabla\times\mathbf{F}$	Eq. (B.1.1.9)
$\nabla\times(\mathbf{F}\times\mathbf{G}) = (\nabla\cdot\mathbf{G})\mathbf{F} - (\nabla\cdot\mathbf{F})\mathbf{G} + (\mathbf{G}\cdot\nabla)\mathbf{F} - (\mathbf{F}\cdot\nabla)\mathbf{G}$	Eq. (B.1.1.10)

The two other possible second-order operations are taking the curl of the curl or the gradient of the divergence of a vector field. It is left as an exercise to show that in rectangular coordinates

$$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2\mathbf{F}$$

Where the Laplacian of a vector is the vector whose rectangular components are the Laplacians of the rectangular components of the original vector. In any coordinate system other than rectangular the Laplacian of a vector is defined by Eq. (B.51).

Another way in which the application of the vector differential operators may be extended is to apply them to various products of two vectors and scalars. The six possible combinations of differential operators and products are tabulated in Table B.1. These identities may be readily verified in rectangular coordinates, which is sufficient to assure their validity in any coordinate system. A derivative of a function, can be calculated by repeated applications of the identities in Table B.1, which is therefore exhaustive. The formulas can be easily remembered from the rules of vector algebra and ordinary differentiation. The only ambiguity could be in Eq. (B.1.1.6) where $\mathbf{F}\cdot\nabla$ occurs (not $\nabla\cdot\mathbf{F}$).

Some particular types of functions come up often enough in electromagnetic theory or heat transfer that is worth noting their various derivatives now. For the function $\mathbf{F} = \mathbf{r}$,

$$\begin{aligned} &\nabla\cdot\mathbf{r} = 3 \\ &\nabla\times\mathbf{r} = 0 \\ &(\nabla\cdot\mathbf{G})\mathbf{r} = \mathbf{G} \\ &\nabla^2\mathbf{r} = 0 \end{aligned} \tag{B.50}$$

For a function that depends only on the distance $\mathrm{r} = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$,

$$\varphi(r) \quad \text{or} \quad \mathbf{F}(r): \quad \nabla = \frac{\mathbf{r}}{r}\frac{d}{dr} \tag{B.51}$$

For a function that depends on the scalar argument $\mathrm{A}\cdot\mathbf{r}$, where $\mathbf{A}$ is a constant vector,

$$(\mathrm{A}\cdot\mathbf{r}) \quad \text{or} \quad \mathbf{F}(\mathrm{A}\cdot\mathbf{r}): \quad \nabla = \mathrm{A}\frac{d}{d(\mathrm{A}\cdot\mathbf{r})} \tag{B.52}$$

For a function that depends on the argument $\mathbf{R} = \mathbf{r} - \mathbf{r}'$ where $\mathbf{r}'$ is treated as a constant,

$$\nabla = \nabla_R \tag{B.53}$$

$$\nabla_R = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$$

where $\mathbf{R} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}$. If $\mathbf{r}$ is treated as constant instead,

$$\nabla = -\nabla' \tag{B.54}$$

where

$$\nabla' = \mathbf{i}\frac{\partial}{\partial x'} + \mathbf{j}\frac{\partial}{\partial y'} + \mathbf{k}\frac{\partial}{\partial z'}$$

There are several possibilities for the extension of the divergence theorem and of Stock's theorem. The most interesting of these is Green's theorem, which is

$$\int_V \left(\psi\nabla^2\varphi - \varphi\nabla^2\psi\right)dv = \oint_S (\psi\nabla\varphi - \varphi\nabla\psi)\cdot\mathbf{n}\,da \tag{B.55}$$

This theorem follows from the application of the divergence theorem to the vector

$$\mathbf{F} = \psi\nabla\varphi - \varphi\nabla\psi$$

Using this $\mathbf{F}$ in the divergence theorem, we obtain

$$\int_V^{\nabla} [\psi\nabla\varphi - \varphi\nabla\psi] = \oint_S (\psi\nabla\varphi - \varphi\nabla\psi)\cdot\mathbf{n}\,da \tag{B.56}$$

Using the identity (Table 1.1) for the divergence of scalar times a vector gives

Table B.2 Vector integral theorems

$\int_S \mathbf{n}\times\nabla\varphi da = \oint_S \varphi d\mathbf{l}$	Eq. (B.1.2.1)
$\int_V \nabla\varphi dv = \oint_S \varphi\mathbf{n} da$	Eq. (B.1.2.2)
$\int_V \nabla\times\mathbf{F} dv = \oint_S \mathbf{n}\times\mathbf{F} da$	Eq. (B.1.2.3)
$\int_V (\nabla\cdot\mathbf{G} + \mathbf{G}\cdot\nabla)\mathbf{F} dv = \oint_S \mathbf{F}(\mathbf{G}\cdot\mathbf{n}) da$	Eq. (B.1.2.4)

$$\nabla\cdot(\varphi\nabla\psi) - \nabla\cdot(\varphi\nabla\psi) = \left(\psi\nabla^2\varphi\right) - \varphi\nabla^2\psi \tag{B.57}$$

Combining Eqs. (B.56) and (B.57) yields Green's theorem. Some other integral theorems are in Table B.2.

This concludes our brief discussion of vector analysis. In the interests of brevity, the proofs of many results have been relegated to the problems. No attempt has been made to achieve a high degree of rigor. The approach has been utilitarian: What we will need, we have developed; everything else has been omitted.

B.9 Summary

Three different types of vector differentiation can be expressed by the **vector differential operator del**, ∇, namely, divergence, and curl:

$$\nabla\varphi = \mathbf{i}\frac{\partial\varphi}{\partial x} + \mathbf{j}\frac{\partial\varphi}{\partial y} + \mathbf{k}\frac{\partial\varphi}{\partial z}$$

$$\nabla\cdot\mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

$$\nabla\times\mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix}$$

Del is a linear operator. Repeated applications of it, or its application to products of functions, produce formulas that can be derived in rectangular coordinates, but are independent of the coordinate system. They can be remembered by the rules of vector algebra and ordinary differential. The derivatives of a few special functions are worth committing to memory. The most important **integral theorems** about the derivatives are

$$\int_{a(C)}^{b} \nabla\varphi \cdot \, d\mathbf{l} = \varphi|_a^b$$

$$\int_S \nabla \times \mathbf{F} \cdot \mathbf{n}\, da = \oint_C \mathbf{F} \cdot d\mathbf{l} \quad \text{(Stokes's theorem)}$$

$$\int_V \nabla \cdot \mathbf{F}\, dv = \oint_S \mathbf{F} \cdot \mathbf{n}\, da \qquad \text{(Divergence theorem)}$$

These are generalizations of the functional theorem of calculus.

B.10 Examples and Solved Problems

Example B.1 Let $\vec{C} = \vec{A} - \vec{B}$ using diagram below, and calculate the dot product of $\vec{C}$ with itself. Using dot product expression Eq. (B.7)

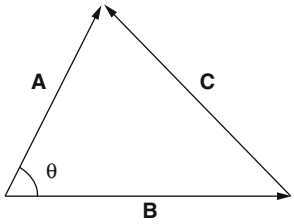


Diagram of Problem B.1

Solution

$$\vec{C} \cdot \vec{C} = \left(\vec{A} - \vec{B}\right) \cdot \vec{A} - \vec{B} = \vec{A} \cdot \vec{B} - \vec{A} \cdot \vec{B} + \vec{B} + \vec{B} \cdot \vec{B}$$

or

$$C^2 = A^2 + B^2 - 2AB\cos\theta$$

This is the **law of cosines**.

Example B.2 Using the definition of dot product and cross product along with the diagram below,

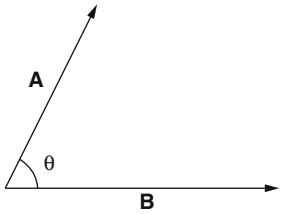


Diagram.A of Problem B.2

Dot Product is defined as;

$\vec{A} \cdot \vec{B} = AB\cos\theta$ The result is ***SCALAR*** form.

Cross Product is defined as;

$\vec{A} \times \vec{B} = AB\sin\theta\hat{n}$ The result is a ***VECTOR*** form

where $\hat{n}$ is a unit vector (vector of length 1) pointing perpendicular to the plane of $\vec{A}$ and B.

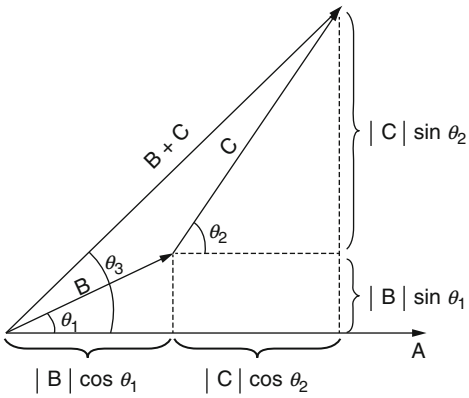


Diagram.B of Problem B.2

Show that the dot product and cross product are distributive,

(a) When the three vectors are coplanar;
(b) In the general case.

Solution

(a) From the diagram below, we have: $\left|\vec{B}+\vec{C}\right|\cos\theta_3 = \left|\vec{B}\right|\cos\theta_1 + \left|\vec{C}\right|\cos\theta_2$.

Multiply by $\left|\vec{A}\right|$ we have then, $\left|\vec{A}\right|\left|\vec{B}+\vec{C}\right|\cos\theta_3 = \left|\vec{A}\right|\left|\vec{B}\right|\cos\theta_1 + \left|\vec{A}\right|\left|\vec{C}\right|\cos\theta_2$

So $\vec{A}\cdot\left(\vec{B}+\vec{C}\right) = \vec{A}\cdot\vec{B} + \vec{A}\cdot\vec{C}$ (Dot product is distributive).

Similarly: $\left|\vec{B}+\vec{C}\right|\sin\theta_3 = \left|\vec{B}\right|\sin\theta_1 + \left|\vec{C}\right|\sin\theta_2$. Multiply by $\left|\vec{A}\right|\hat{n}$ we have then, $\left|\vec{A}\right|\left|\vec{B}+\vec{C}\right|\sin\theta_3\hat{n} = \left|\vec{A}\right|\left|\vec{B}\right|\sin\theta_1\hat{n} + \left|\vec{A}\right|\left|\vec{C}\right|\sin\theta_2\hat{n}$.

If $\hat{n}$ is the unit vector pointing out of the page, it follows that $\vec{A}\times\left(\vec{B}+\vec{C}\right) = \left(\vec{A}\times\vec{B}\right) + \left(\vec{A}\times\vec{C}\right)$ (Cross product is distributive)

(b) For the general case, see *G. E. Hay's Vector and Tensor Analysis*, Chap. 1, Sects. 7 (dot product) and 8 (cross product).

Example B.3 Is the cross product associative?

$$\left(\vec{A} \times \vec{B}\right) \times \vec{C} \neq \vec{A} \times \left(\vec{B} + \vec{C}\right)$$

Solution The triple cross-product is *not* in general associative. For example, suppose $\vec{A} = \vec{B}$ and $\vec{C}$ is perpendicular to $\vec{A}$, as in the diagram. Then $\left(\vec{B} \times \vec{C}\right)$ points out-of-the-page, and $\vec{A} \times \left(\vec{B} + \vec{C}\right)$ points *down*, and has magnitude *ABC*. But $\left(\vec{A} \times \vec{B}\right) = 0$, so $\left(\vec{A} \times \vec{B}\right) \times \vec{C} = 0 \neq \vec{A} \times \left(\vec{B} \times \vec{C}\right)$

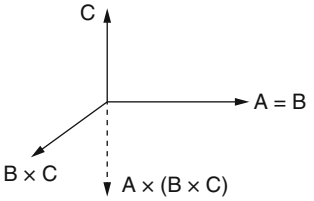


Diagram of Problem B.3

Example B.4 Find the angle between the face diagonals of a cube, using following diagram

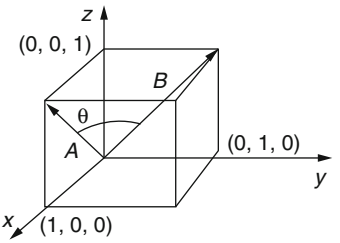


Diagram of Problem B.3

Solution We might as well use a cube of side I, and place it as shown in above figure, with one corner at the origin. The face diagonals $\vec{A}$ and $\vec{B}$ are;

$$\vec{A} = 1\hat{x} + 0\hat{y} + 1\hat{z} \quad \vec{B} = 0\hat{x} + 1\hat{y} + 1\hat{z}$$

So, in component form,

$$\vec{A} \cdot \vec{B} = 1.0 + 0.1 + 1.1$$

On the other hand, in "abstract" form,

$$\vec{A} \cdot \vec{B} = AB\cos\theta = \sqrt{2}\sqrt{2}\cos\theta = 2\cos\theta$$

Therefore;

$$\cos\theta = 1/2 \text{ or } \theta = 60$$

Of course, you can get the answer more easily by drawing in a diagonal across the top of the cube, completing the equilateral triangle. But in cases where the geometry is not so simple, this device of comparing the abstract and compohent forms of the dot product can be a very efficient means of finding angles.

Example B.5 Find the gradient of $r = \sqrt{x^2 + y^2 + z^2}$ (the magnitude of the position vector).

Solution

$$\begin{aligned}\nabla r &= \frac{\partial r}{\partial x}\hat{x} + \frac{\partial r}{\partial y}\hat{y} + \frac{\partial r}{\partial z}\hat{z} \\ &= \frac{1}{2}\frac{2x}{\sqrt{x^2 + y^2 + z^2}}\hat{x} + \frac{1}{2}\frac{2y}{\sqrt{x^2 + y^2 + z^2}}\hat{y} + \frac{1}{2}\frac{2z}{\sqrt{x^2 + y^2 + z^2}}\hat{z} \\ &= \frac{x\hat{x} + y\hat{y} + z\hat{z}}{\sqrt{x^2 + y^2 + z^2}} = \frac{\vec{r}}{r} = \hat{r}\end{aligned}$$

Does this make sense? Well, it says that the distance from the origin increases most rapidly in the radial direction, and that its rate of increase in that direction is 1.

Example B.6 Suppose the vectors are in form of $\vec{V}_a = \vec{r} = x\hat{x} + y\hat{y} + z\hat{z}$, $\vec{V}_b = \vec{r} = \hat{z}$, and $\vec{V}_c = \vec{r} = z\hat{z}$. Calculate their divergences.

Solution

$$\nabla \cdot \vec{V}_a = \frac{\partial}{\partial x}(x) + \frac{\partial}{\partial y}(y) + \frac{\partial}{\partial z}(z) = 1 + 1 + 1 = 3$$

As anticipated, this function has a positive divergence.

$$\nabla \cdot \vec{V}_b = \frac{\partial}{\partial x}(0) + \frac{\partial}{\partial y}(0) + \frac{\partial}{\partial z}(1) = 0 + 0 + 0 = 0$$

$$\nabla \cdot \vec{V}_c = \frac{\partial}{\partial x}(0) + \frac{\partial}{\partial y}(0) + \frac{\partial}{\partial z}(z) = 0 + 0 + 1 = 1$$

Example B.7 Suppose the vectors are in form of $\vec{V}_a = \vec{r} = -y\hat{x} + x\hat{y}$, and $\vec{V}_b = \vec{r} = x\hat{y}$. Calculate their curls.

Solution

$$\nabla \mathrm{X}\, \vec{V}_a = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ -y & x & 0 \end{vmatrix} = 2\hat{z}$$

and

$$\nabla \mathrm{X}\, \vec{V}_b = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ 0 & x & 0 \end{vmatrix} = \hat{z}$$

As expected, these curls point in the + z-direction. (Incidentally, they both have zero divergence.)

Appendix C
Short Course in Ordinary and Partial Differential Equations

The laws of nature are written in the language of Ordinary and Partial Differential Equations. Therefore, these equations arise as models in virtually all branches of science and technology. Our goal in this section is to help you to understand what this vast subject is about. The section is an introduction to the field. We assume only that you are familiar with basic calculus and elementary linear algebra. Introductory courses in differential and partial differential equations are given all over the world in various forms. The traditional approaches to the subjects are to introduce a number of analytical techniques, enabling the student to derive exact solutions of some simplified problems.

We introduce analytical and computational techniques in this section to some degree. The main reason for doing this is that the computer, developed to assist scientists in solving differential and partial differential equations, has become commonly available and is currently used in all practical applications of partial differential equations. Therefore, a modern introduction to this topic must focus on methods suitable for computers. But these methods often rely on deep analytical insight into the equations. We must therefore take great care not to throw away basic analytical methods but seek a sound balance between analytical and computational techniques.

C.1 Differential Equations

The theory of differential equations is quite developed and the methods used to study them vary significantly with the type of the equation. Many physical situations are represented mathematically by equations that involve a variable and its rate of change. For example, mathematical model of the natural cooling of any object relates the temperature T of the object to its rate of change. In symbolic form we would write this as

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

$$\frac{dT}{dt} = \alpha T + \beta \tag{C.1}$$

Such an equation cannot be integrated directly with respect to the time variable t using the techniques of any normal integration, because the right-hand side involves the function T. We need a new set of techniques to solve equations of this type. An equation which involves the dependent variable T and its derivative is called an *ordinary differential equation* and in the next few section we investigate methods of solving them.

C.1.1 Definition

- An ordinary differential equation (ODE) is a differential equation in which the unknown function (also known as the dependent variable) is a function of a single independent variable. In the simplest form, the unknown function is a real or complex valued function, but more generally, it may be vector-valued or matrix-valued: this corresponds to considering a system of ordinary differential equations for a single function. Ordinary differential equations are further classified according to the order of the highest derivative of the dependent variable with respect to the independent variable appearing in the equation. The most important cases for applications are first-order and second-order differential equations. In the classical literature also distinction is made between differential equations explicitly solved with respect to the highest derivative and differential equations in an implicit form.
- A partial differential equation (PDE) is a differential equation in which the unknown function is a function of multiple independent variables and the equation involves its partial derivatives. The order is defined similarly to the case of ordinary differential equations, but further classification into elliptic, hyperbolic, and parabolic equations, especially for second-order linear equations, is of utmost importance. Some partial differential equations do not fall into any of these categories over the whole domain of the independent variables and they are said to be of mixed type. (See Sect. 2 of this appendix)

Both ordinary and partial differential equations are broadly classified as linear and nonlinear. A differential equation is linear if the unknown function and its derivatives appear to the power 1 (products are not allowed) and nonlinear otherwise. The characteristic property of linear equations is that their solutions form an affine subspace of an appropriate function space, which results in much more developed theory of linear differential equations. Homogeneous linear differential equations are a further subclass for which the space of solutions is a linear subspace, i.e., the sum of any set of solutions or multiples of solutions is also a solution. The coefficients of the unknown function and its derivatives in a linear differential

equation are allowed to be (known) functions of the independent variable or variables; if these coefficients are constants, then one speaks of a constant coefficient linear differential equation.

There are very few methods of explicitly solving nonlinear differential equations; those that are known typically depend on the equation having particular symmetries. Nonlinear differential equations can exhibit very complicated behavior over extended time intervals, characteristic of chaos. Even the fundamental questions of existence, uniqueness, and extendability of solutions for nonlinear differential equations, and well-posedness of initial and boundary value problems for nonlinear PDEs are hard problems and their resolution in special cases is considered to be a significant advance in the mathematical theory (cf. Navier–Stokes existence and smoothness).

Linear differential equations frequently appear as approximations to nonlinear equations. These approximations are only valid under restricted conditions. For example, the harmonic oscillator equation is an approximation to the nonlinear pendulum equation that is valid for small amplitude oscillations (see below).

C.1.2 Types of ODE and PDE

In the first group of examples, let u be an unknown function of x, and c and ω are known constants.

1. Inhomogeneous first-order linear constant coefficient ordinary differential equation:

$$\frac{du}{dx} = cu + x^2$$

2. Homogeneous second-order linear ordinary differential equation:

$$\frac{d^2u}{dx^2} - x\frac{du}{dx} + u = 0$$

3. Homogeneous second-order linear constant coefficient ordinary differential equation describing the *harmonic oscillator*:

$$\frac{d^2u}{dx^2} + \omega^2 u = 0$$

4. First-order nonlinear ordinary differential equation:

$$\frac{du}{dx} = u^2 + 1$$

5. Second-order nonlinear ordinary differential equation describing the motion of a pendulum of length L:

$$g\frac{d^2u}{dx^2} + L\ \sin u = 0$$

In the next group of examples, the unknown function u depends on two variables x and t or x and y.

6. Homogeneous first-order linear partial differential equation:

$$\frac{\partial u}{\partial t} + t\frac{\partial u}{\partial x} = 0$$

7. Homogeneous second-order linear constant coefficient partial differential equation of elliptic type, the *Laplace equation*:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

8. Third-order nonlinear partial differential equation, the *Korteweg–de Vries equation*:

$$\frac{\partial u}{\partial t} = 6u\frac{\partial u}{\partial x} - \frac{\partial^3 u}{\partial x^3}$$

C.1.3 Classification of Differential Equations

Before we begin we need to introduce a simple classification of differential equations which will let us increase the complexity of the problems we consider in a systematic way. In order to talk about Differential Equation, we shall classify them by *Type*, *Order*, and *Linearity*.

C.1.4 Classification of Differential Equations by Type

If an equation contains only ordinary derivatives of one or more dependent variables with respect to a single independent variable, it is said to be an **Ordinary Differential Equation (ODE)**. For example

$$\frac{dy}{dx} + 5y = e^x \quad \frac{d^2y}{dx^2} - \frac{dy}{dx} + 6y = 0 \quad \frac{dx}{dt} + \frac{dy}{dt} = 2x + y \tag{C.2}$$

are ordinary differential equations.

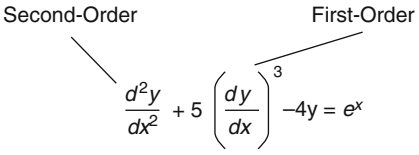
Note that if an equation involving the partial derivative of one or more dependent variables of two or more independent variables is called a **Partial Differential Equation (PDE)**. For example

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2} - 2\frac{\partial u}{\partial t} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{C.3}$$

are partial differential equations. In general the nth derivative is written as $d^n y/dx^n$ or $y^{(n)}$. In most books the **Leibniz Notation** such as $dy/dx, d^2y/dx^2, d^3y/dx^3, \ldots$. over the **Prime Notation** such as $y', y'', y''', \ldots$. Also sometime you will the **Newton's Dot Notation** as well such as, $\dot{y}, \ddot{y}, \dddot{y}, \ldots$.

C.1.5 Classification of Differential Equations by Order

The **Order of a Differential Equation** (either ODE or PDE) is the order of the highest derivative in the equation. For example,



is a second-order ordinary differential equation. First-Order ordinary differential equations are occasionally written in differential form $M(x, y)dx + N(x, y)dy = 0$. For example, if we assume that y denotes the dependent variable in $(y - x)dx + 4xdy = 0$, then $y' = dy/dx$, and so by dividing by the differential dx we get the alternative form $4xy' + y = x$.

In general we can express an nth-order ordinary differential equation in one dependent variable by the general for of;

$$F\left(x, y, y', y'', \ldots\ldots, y^{(n)}\right) = 0 \tag{C.4}$$

Where F is a real-valued function of $n + 2$ variables, $x, y, y', y'', \ldots\ldots, y^{(n)}$, and where $y^{(n)} = d^n y/dx^n$. We are making the assumption both from practical and mathematical point of view, it is possible to solve an ordinary differential equation form of Eq. (C.4) uniquely for the highest derivative $y^{(n)}$ in terms of the remaining $n + 1$ variables. The differential equation

$$\frac{d^n y}{dx^n} = f\left(x, y, y', y'', \ldots\ldots, y^{(n)}\right) \tag{C.5}$$

where f is a real-valued continuous function, is referred to as the **Normal Form** of Eq. (C.4). Thus when it suits our purposes, we shall use the normal forms of

$$\frac{dy}{dx} = f(x, y) \quad \text{and} \quad \frac{d^2 y}{dx^2} = f\left(x, y, y'\right)$$

to represent general first- and second-order ordinary differential equations. For example, the normal form of the first-order equation $4xy' + y = x$ is $y' = (x - y)/4x$.

C.1.6 Classification of Differential Equations by Linearity

An nth-order ordinary differential Eq. (C.4) is said to be **Linear** if F is linear in y, $y', y'', \ldots\ldots, y^{(n)}$. This means that an nth order ODE is linear when Eq. (C.4) is

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \ldots. + a_1(x)y' + a_0(x)y - g(x) = 0$$

or we can write as follows:

$$a_n(x)\frac{d^n y}{dx^n} + a_{n-1}(x)\frac{d^{(n-1)} y}{dx^{(n-1)}} + \ldots. + a_1(x)\frac{dy}{dx} + a_0(x)y - g(x) = 0 \tag{C.6}$$

From Eq. (C.6) we see the characteristic two properties of a linear differential equation:

1. The dependent variable and all its derivatives are of the first degree—that is, the power of each term involving y is 1.
2. Each coefficient depends at most on the independent variable x.

The equations

$$(y - x)dx + 4xdy = 0 \quad y'' - 2y' + y = 0 \quad \text{and} \quad \frac{d^3 y}{dx^3} + x\frac{dy}{dx} - 5y = e^x$$

are, in turn, linear first-, second-, and third-order ordinary differential equations. Note that the first equation in above is linear in the variable y by writing it in the alternative for $4xy' + y = x$. A **Nonlinear** ordinary differential equation is simply one that is not linear. Nonlinear functions of the dependent variable or its derivatives, such as $\sin\ y$ or $e^{y'}$, cannot appear in a linear equation. Therefore,

nonlinear term: coefficient depends on y ↓ $(1-y)y' + 2y = e^x$

nonlinear term: coefficient function on y ↓ $\frac{d^2y}{dx^2} + \sin y = 0$

and

nonlinear term: power not 1 ↓ $\frac{d^4y}{dx^4} + y^2 = 0$

are examples of nonlinear first-, second-, and forth-order ordinary differential equations, respectively.

C.1.7 Initial Value Problems (IVP)

We are often interested in solving a differential equation subject to prescribed side conditions—conditions that are imposed on the unknown solution $y = y(x)$ or its derivatives. On some interval I containing x_0, the problem

$$\begin{aligned} &\textit{Solve}: \quad \frac{d^n y}{dx^n} = f\left(x, y, y', \ldots, y^{(n-1)}\right) \\ &\textit{Subject}: \quad y(x_0) = y_0 \quad y'(x_0) = y_1, \ \ldots \ y^{(n-1)}(x_0) = y_{n-1} \end{aligned} \tag{C.7}$$

where $y_0, y_1, \ldots y_{n-1}$ are arbitrary specified real constants, is called an **Initial Value Problem (IVP)**. The values of $y(x)$ and its first $n-1$ derivatives at a single point $x_0 : y(x_0) = y_0, y'(x_0) = y_1, \ \ldots \ y^{(n-1)}(x_0) = y_{n-1}$ are called **Initial Conditions**.

First- and second-Order IVPs: The problem given in Eq. (C.7) is also called and **nth-order initial value problem**. For example

$$\begin{aligned} &\textbf{\textit{Solve:}} \quad \frac{dy}{dx} = f(x, y) \\ &\textbf{\textit{Subject to:}} \quad y(x_0) = y_0 \end{aligned} \tag{C.8}$$

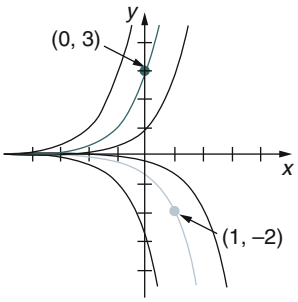
and

$$\begin{aligned} &\textbf{\textit{Solve:}} \quad \frac{d^2y}{dx^2} = f\left(x, y, y'\right) \\ &\textbf{\textit{Subject to:}} \quad y(x_0) = y_0, \quad y'(x_0) = y_1 \end{aligned} \tag{C.9}$$

are first- and second-order initial-value problem, respectively [5].

Example 1: First-Order IVP It is readily verified that $y = ce^x$ is a one-parameter family of solutions of the simple first-order equation $y' = y$ on the interval $(-\infty, \infty)$. If we specify an initial condition, say, $y(0) = 3$, then substituting $x = 0$, $y = 3$, in the family determines the constant $3 = ce^x = c$. Thus the function $y = ce^x$ is a solution of the initial-value problem.

$$y' = y \quad y(0) = 3$$



Now if we demand that a solution of the differential equation pass through the point $(1, -2)$ rather that $(0, 3)$, then $y(1) = -2$ will yield $-2 = ce$ or $c = -2e^{-1}$. The function $y = -2e^{x-1}$ is a solution of the initial value problem.

Example 2: Second-Order IVP Before we do this type of problem we show that the function $x = c_1 \cos 4t$ and $x = c_2 \sin 4t$, where c_1 and c_2 are arbitrary constants or parameters are both solutions of the linear differential equation $x'' + 16x = 0$.

For $x = c_1 \cos 4t$ the first two derivatives with respect to t are $x' = -4c_1 \sin 4t$ and $x'' = -16c_1 \cos 4t$. Substituting x'' and x then gives;

$$x'' + 16x = -16c_1 \cos 4t + 16c_1 \cos 4t = 0$$

Similarly for $x = c_2 \sin 4t$ we have $x'' = -16c_2 \sin 4t$ and so

$$x'' + 16x = -16c_2 \sin 4t + c_2 \sin 4t = 0$$

Finally is straightforward to verify that the linear combination of solutions or the two-parameter family $x = c_1 \cos 4t + c_2 \sin 4t$ is also a solution of the differential equation. Now we turn our task to Second-Order IVP in hand problem.

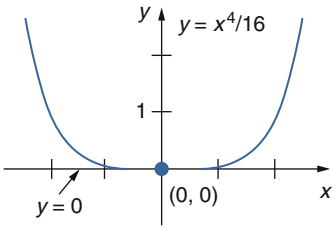
In this case we want to find a solution of the initial-value problem

$$x'' + 16x = 0 \quad x\left(\frac{\pi}{2}\right) = -2 \quad \text{and} \quad x'\left(\frac{\pi}{2}\right) = 1 \tag{C.10}$$

We first apply $x(\pi/2) = -2$ to the given family of solutions: $c_1 \cos 2\pi + c_2 \sin 2\pi = -2$. Since $\cos 2\pi = 1$ and $\sin 2\pi = 0$, we find that $c_1 = -2$. We next apply $x'(\pi/2) = 1$ to the one-parameter family $x(t) = -2 \cos 4t + c_2 \sin 4t$. Differentiating and then setting $t = \pi/2$ and $x'(\pi/2) = 1$ 8 $\sin 2\pi + 4c_2 \cos 2\pi = 1$, from which we see that $c_2 = \frac{1}{4}$. Hence $x = -2 \cos 4t + \frac{1}{4} \sin 4t$ is a solution of (I).

Example 3: An IVP Can Have Several Solutions Each of the functions $y = 0$ and $y = \frac{1}{16}x^4$ satisfies the differential equation $dy/dx = xy^{1/2}$ and the initial condition $y(0) = 0$, and so the initial-value problem

$$\frac{dy}{dx} = xy^{1/2} \quad \text{and} \quad y(0) = 0$$



has at least two solutions. As illustrated in the figure here, the graphs of both functions pass through the same point $(0, 0)$.

Within safe confines one can be fairly confident that *most* differential equations will have solutions and that solutions of initial-value problem will *probably* be unique.

C.1.8 Linear First-Order Differential Equations

A differential equation is said to be linear when it is of the first degree in the dependent variable and all its derivatives. When $n = 1$ in Eq. (C.6), we obtain a linear first-order differential equation.

Definition 1: Separable Equation
A first-order differential equation of the form

$$\begin{aligned} \frac{dy}{dx} &= g(x)h(y) \\ \frac{1}{h(y)}dy &= g(x)dx \\ p(y)dy &= g(x)dx \end{aligned} \tag{C.11}$$

is said to be a **Separable** or have **Separable Variables**. Note that we have denoted the $\frac{1}{h(y)}$ by $p(y)$.

For example the equations

$$\frac{dy}{dx} = y^2 x e^{3x+4y} \quad \text{and} \quad \frac{dy}{dx} = y + \sin\ x$$

are separable and non separable, respectively. In the first equation we can factor $f(x, y) = y^2 x e^{3x+4y}$ as

$$f(x,y) = \begin{matrix} g(x) & h(y) \\ \downarrow & \downarrow \\ (xe^{3x}) & (y^2 e^{4y}) \end{matrix}$$

but in the second equation there is no way of expressing $f(x, y) = y + \sin\ x$ as a product of a function of x times a function of y.

Definition 2: Linear Equation
A first-order differential equation of the form

$$a_1(x)\frac{dy}{dx} + a_0(x)y = g(x) \tag{C.12}$$

is said to be a **linear equation**.

When $g(x) = 0$, the linear equation is said to be **homogenous**; otherwise, it is **nonhomogeneous**.

By dividing both side of Eq. (C.8) by the lead coefficient $a_1(x)$ we obtain a more useful form that is known as **Standard Form** of a linear equation as follows:

$$\frac{dy}{dx} + P(x)y = f(x) \tag{C.13}$$

Where $P(x) = \dfrac{a_0(x)}{a_1(x)}$ and $f(x) = \dfrac{g(x)}{a_1(x)}$ are respectively. We now seek a solution for Eq. (C.9) on an interval I for which both function $P(x)$ and $f(x)$ are continuous. In order to solve Eq. (C.9) we note a **Property** for it that its solution is the **Sum** of the two solutions: $y = y_c + y_p$, where y_c is a **Complementary** solution of the associated homogeneous part of Eq. (C.9) in the following form

$$\frac{dy}{dx} + P(x)y = 0 \tag{C.14}$$

while y_p is the **Particular** solution of the nonhomogeneous part. To see this, please observe that [5].

$$\frac{d}{dx}\left[y_c + y_p\right] + P(x)\left[y_c + y_p\right] = \underbrace{\left[\frac{dy_c}{dx} + P(x)y_c\right]}_{0} + \underbrace{\left[\frac{dy_c}{dx} + P(x)y_1\right]}_{\left(f(x)\right.} = f(x)$$

Now the homogeneous Eq. (C.14) is also separable. This fact enables us to find y_c by writing Eq. (C.14) as if we multiply both sides by $\dfrac{dx}{y}$,

$$\frac{dx}{y} + P(x)dx = 0$$

and integrating. Solving for y gives $y_c = ce^{-\int P(x)dx}$. For convenience let us write $y_c = cy_1(x)$, where $y_1(x) = e^{-\int P(x)dx}$. The fact that $dy_1/dx + P(x)y_1 = 0$ will be used next to determine y_p.

We can now find a particular solution of Eq. (C.9) by a procedure known as **Variation of Parameters**. The basic idea here is to find a function u so that $y_p = u(x)y_1(x) = u(x)e^{-\int P(x)dx}$ is a solution of Eq. (C.9). In order words, our assumption for y_p is the same as $y_c = cy_1(x)$ except that c is replaced by the "Variable Parameter" u. Substituting $y_p = uy_1$ into Eq. (C.2) gives,

$$\underset{\text{Product Rule}}{\underset{\downarrow}{u\frac{dy_1}{dx}+y_1\frac{du}{dx}+P(x)uy_1=f(x)}} \quad \text{or} \quad \underset{\text{Zero}}{\underset{\downarrow}{u\left[\frac{dy_1}{dx}+P(x)y_1\frac{du}{dx}\right]}}=f(x)$$

so

$$y_1\frac{du}{dx}=f(x)$$

Separating variables and integrating then gives

$$du\frac{f(x)}{y_1(x)}dx \quad \text{and} \quad u=\int\frac{f(x)}{y_1(x)}dx$$

Since $y_1(x)=e^{-\int P(x)dx}$, we see that $1/y_1(x)=e^{\int P(x)dx}$. Therefore;

$$y_p=uy_1\left(\int\frac{f(x)}{y_1(x)}dx\right)e^{-\int P(x)dx}=e^{-\int P(x)dx}\int e^{\int P(x)dx}f(x)dx$$

and

$$y=\underbrace{ce^{-\int P(x)dx}}_{y_c}+\underbrace{e^{-\int P(x)dx}\int e^{\int P(x)dx}f(x)d\,x}_{y_P} \tag{C.15}$$

Here if Eq. (C.9) has a solution, it must be of form Eq. (C.15). Closely, it is a straightforward exercise in differentiation to verify that Eq. (C.4) constitutes a one-parameter family of solution of Eq. (C.9).

We do not need to remember or memorize the formula in Eq. (C.15). However, you should remember the special term

$$e^{\int P(x)dx} \tag{C.16}$$

because it is used in the equivalent but easier way of solving Eq. (C.9). If Eq. (C.15) is multiplied by Eq. (C.16), then we have

$$e^{\int P(x)dx}\, y = c + \int e^{\int P(x)dx} f(x)dx \tag{C.17}$$

and then Eq. (C.17) is differentiated.

$$\frac{d}{dx}\left[e^{\int P(x)dx}\, y\right] = e^{\int P(x)dx} f(x) \tag{C.18}$$

we get

$$e^{\int P(x)dx}\, \frac{dy}{dx} + P(x)e^{\int P(x)dx} = e^{\int P(x)dx} f(x)dx \tag{C.19}$$

Dividing Eq. (C.19) by $e^{\int P(x)dx}$ gives Eq. (C.9) [5].

C.1.8.1 Method of Solving a Linear First-Oder Equation

The recommended method of solving Eq. (C.9) actually consists of Eqs. (C.17–C.19) worked in reverse order. Because we can solve Eq. (C.9) by integrating after multiplication by $e^{\int P(x)dx}$, we call this function an **Integral Factor** for the differential equation. In brief summary the followings are the steps one needs to take in order to solve such first-order differential equations;

1. Put a linear equation of form Eq. (C.8) into the standard form of Eq. (C.9).
2. From the standard form identify $P(x)$ and then the integrating factor $e^{\int P(x)dx}$.
3. Multiply the standard form of the equation by the integrating factor. The left-hand side of the resulting equation is automatically the derivative of the integrating factor and y:

$$\frac{d}{dx}\left[e^{\int P(x)dx}\, y\right] = e^{\int P(x)dx} f(x)$$

4. Integrate both sided of this last equation.

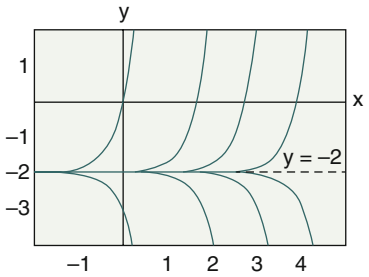


Fig. C.1 Some solution of $y' - 3y = 6$ [5]

Example 1: Solving a Homogeneous Linear Differential Equation Solve $\dfrac{dy}{dx} - 3y = 0$

Solution: This linear equation can be solved by separation of variables. Alternatively, since the equation is already in the standard form Eq. (C.9), we see that $P(x) = -3$ and so the integrating factor is $e^{\int(-3)dx} = e^{-3x}$. We multiply the equation by this factor and recognize that

$$e^{3x}\frac{dy}{dx} - 3e^{3x}y = 0 \quad \text{is the same as} \quad \frac{d}{dx}\left[e^{-3x}y\right] = 0$$

Integrating both sides of the last equation gives $e^{3x}y = c$. Solving for y gives us the explicit solution $y = ce^{3x}$, $-\infty < x < \infty$ (Fig. C.1).

Example 2: Solving a Nonhomogeneous Linear Differential Equation Solve $\dfrac{dy}{dx} - 3y = 6$

Solution: The associated homogeneous equation for this DE was solved in Example 1 in above. Again the equation is already in the standard form Eq. (C.9), and the integrating factor I still $e^{\int(-3)dx} = e^{-3x}$. This time multiplying the given equation by this factor gives

$$e^{3x}\frac{dy}{dx} - 3e^{3x}y = 6e^{-3x} \quad \text{which is the same as} \quad \frac{d}{dx}\left[e^{-3x}y\right]6e^{-3x}$$

Integrating both sides of the last equation gives $e^{-3x}y = -2e^{-3x} + c$ or $y = -2 + ce^{3x}$, $-\infty < x < \infty$.

The final solution in Example 2 is the sum of two solution $y = y_c + y_p$, where $y_c = ce^{3x}$ is the solution of the homogeneous equation in Example 1 and $y_p = -2$ is a particular solution of the nonhomogeneous equation $y' - 3y = 6$. You need not be concerned about whether a linear first-order equation is homogeneous or

nonhomogeneous; when you follow the solution procedure outlined above, a solution of a nonhomogeneous equation necessarily turns to be $y = y_c + y_p$.

Example 3: General Solution Solve $x\dfrac{dy}{dx} - 4y = x^6 e^x$

Solution: Dividing by x, we get the standard form

$$\frac{dy}{dx} - \frac{4}{x}y = x^5 e^x \tag{C.20}$$

From this form we identify $P(x) = -4/x$ and $f(x) = x^5 e^x$, further observe that P and f are continuous on . Hence the integrating factor is

$$\begin{array}{c} \text{we can use } \ln x \text{ instead of } \ln|x| \text{ since } x > 0 \\ \downarrow \\ e^{-4\int dx/x} = e^{-4\ln x} = e^{\ln x^{-4}} = x^{-4} \end{array}$$

Here we have used the basic identity $b^{\log_n N} = N,\ N > 0$. Now we multiply Eq. (C.20) by x^{-4} and rewrite,

$$x^{-4}\frac{dy}{dx} - 4x^5 y = xe^x \quad \text{as} \quad \frac{d}{dx}\left[x^{-4}y\right] = xe^x$$

It follows from integration by parts that the general solution defined on the interval $x^{-4}y = xe^x - e^x + c$ is or $y = x^5 e^x - x^4 e^x + cx^4$.

Note: Except in the case when the lead coefficient is 1, the recasting of Eq. (C.8) into the standard form Eq. (C.9) requires division by $a_1(x)$. Values of x for which $a_1(x) = 0$ are called **Singular Points** of the equation. Singular points are potentially troublesome. Specifically, in Eq. (C.9), if $P(x)$ (formed by dividing $a_0(x)$ by $a_1(x)$) is discontinuous at a point, the discontinuity may carry over to solution of the differential equations [5].

Example 4: General Solution Find the general solution of $(x^2 - 9)\dfrac{dy}{dx} + xy = 0$.

Solution: We write the differential equation in standard form

$$\frac{dy}{dx} + \frac{x}{x^2 - 9}y = 0 \tag{C.21}$$

and identify $P(x) = x/(x^2 - 9)$. Although P is continuous on $(-\infty, -3)$, $(-3, 3)$, and $(3, \infty)$, we shall solve the equation on the first and third intervals. On these intervals the integrating factor is

$$e^{\int x dx/(x^2-9)} = e^{\frac{1}{2}\int 2x dx/(x^2-9)} = e^{\frac{1}{2}\ln|x^2-9|} = \sqrt{x^2-9}$$

After multiplying the standard form (C.21) by this factor, we get

$$\frac{d}{dx}\left[\sqrt{x^2-9}\right] = 0$$

Integrating both side it will give $\sqrt{x^2-9}y = c$. Thus for either $x > 3$ or $x < -3$ the general solution of the equation is

$$y = \frac{c}{\sqrt{x^2-9}}$$

Notice that in this example $x = 3$ and $x = -3$ are singular points of the equation and that every function in the general solution $y = c/\sqrt{x^2-9}$ is discontinuous at these points.

C.1.8.2 Functions Defined by Integrals

Some simple functions do not possess anti-derivatives that are elementary function, and integrals of these kinds of functions are called non-elementary. Two such functions are the **error function** and the **complementary error function**. For example from calculus we know that $\int e^{x^2} dx$ and $\int \sin\ x^2 dx$ are non-elementary integrals and in applied mathematics some important function are defined in terms of these non-elementary integrals.

Error Function

Two such functions are the **error function** and the **complementary error function** as follows;

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2} dt \quad \text{and} \quad erfc(x) = \frac{2}{\sqrt{\pi}}\int_0^\infty e^{-t^2} dt$$

Since $\lim_{x\to\infty} erf(x) = (2/\sqrt{\pi})\int_0^\infty e^{-t^2} dt = 1$, it is seen from above relationship that the error function $erf(x)$ and the complementary error function $erfc(x)$ are related by $erf(x) + erfc(x) = 1$ (Fig. C.2).

Fig. C.2 Error Function

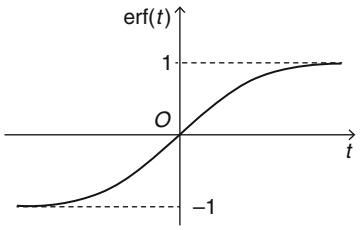
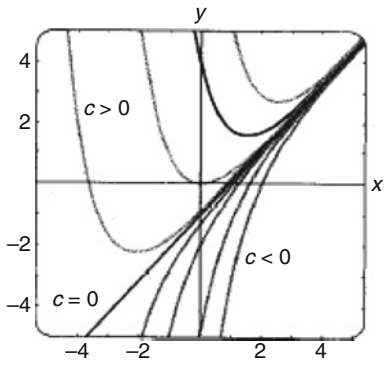


Fig. C.3 Some solutions of $y' + 2xy = 2$



Example: Solve the initial-value problem $\dfrac{dy}{dx} - 2xy = 2,\ y(0) = 1.$

Solution: Since the equation is already in standard form

$$\frac{dy}{dx}\left[e^{-x^2}\right] = 2e^{-x^2} \quad \text{we get} \quad y = 2e^{x^2}\int_0^x e^{-t^2}dt + ce^{x^2} \tag{C.22}$$

$$y = 2e\int_0^x e^{-t^2}dt + e^{x^2} = e^{x^2}\left[1 + \sqrt{\pi}erf(x)\right]$$

The graph of this solution on $(-\infty, \infty)$, shown in Fig. C.3 above among other members of the family defined by (C.22) was obtained with the aid of a compute algebra system such as Mathematica [7].

Dirac Delta Function

Another special function that is convenient to work with is Delta function and is defined as below

$$\delta(t - t_0) = \lim_{a \to \infty} \delta_a(t - t_0)$$

The later expression, which is not a function at all, can be characterized by the two properties

1. $\delta(t - t_0) = \begin{cases} \infty & t = t_0 \\ 0 & t \neq t_0 \end{cases}$
2. $\int_0^\infty \delta(t - t_0)dt = 1$

The unit impulse $\delta(t - t_0)$ is called the **Dirac Delta Function**. It is possible to obtain the Laplace transform of Dirac delta function by the formal assumption that $\pounds[\delta(t - t_0)] = \lim_{a \to 0} \pounds[\delta_a(t - t_0)]$ [7]. See Sect. 2.1 of Appendix E.

Beta Function

In mathematics, the beta function, also called the Euler integral of the first kind, is a special function defined by

$$\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1 - t)^{y-1}dt$$

for $\mathrm{Re}(x), \mathrm{Re}(y) > 0$.

The beta function was studied by Euler and Legendre and was given its name by Jacques Binet; its symbol B is a Greek capital β.

Gamma Function

Euler's integral definition of the **Gamma Function** is as follow

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \tag{C.23}$$

Convergence of the integral requires that $x - 1 > -1$ or $x > 0$. The recurrence relation

$$\Gamma(x + 1) = x\Gamma(x) \tag{C.24}$$

Equation C.24 can be obtained from Eq. (C.23) with integration by parts. Now when $x = 1$, $\Gamma(1) = \int_0^\infty e^{-t}dt = 1$, and thus Eq. (C.24) gives

$$\Gamma(2) = 1\Gamma(1) = 1$$

$$\Gamma(3) = 2\Gamma(2) = 2.1$$

$$\Gamma(4) = 3\Gamma(3) = 3.2.1$$

And so on. In this case manner it is seen that when n is a positive integer then we have,

$$\Gamma(n+1) = n!$$

For this reason the gamma function is often called the **Generalized Factorial Function**.

Although the integral form Eq. (C.23) does not converges for $x < 0$, it can be shown by means of alternative definitions that the gamma function is defined for all real and complex numbers *except* $x = -n$, $n = 0, 1, 2, \cdots$. As a consequence, Eq. (C.24) is actually valid for $x \neq -n$. The graph of $\Gamma(x)$,considered as a function of a real variable x, is as given in Fig. C.4.

Bessel Function

The two differential equations occur frequently in advanced studies in applied mathematics, physics, and engineering are the as follows;

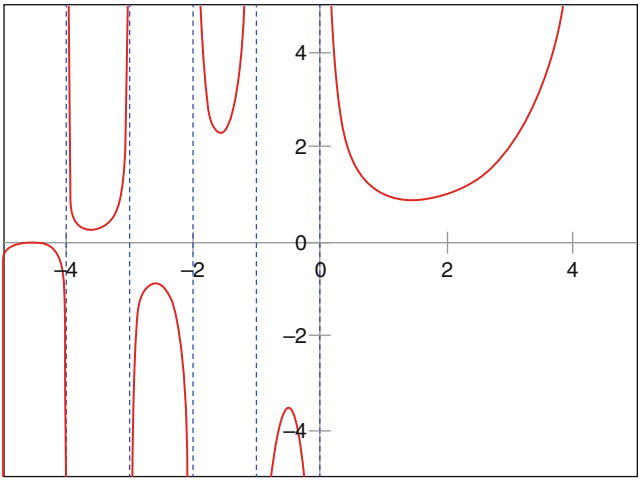


Fig. C.4 The Gamma function along part of the real axis

$$x^2y'' + xy' + \left(x^2 - v^2\right)y = 0 \tag{C.25}$$

$$\left(1 - x^2\right)y'' + 2xy' + n(n+1)y = 0 \tag{C.26}$$

They are called **Bessel's Equation** and **Legendre's Equation** respectively. In solving Eq. (C.25) we shall assume $v \geq 0$, whereas in Eq. (C.26) we shall consider only the case when n is a non-negative integer. Since we seek series solutions of each equation about $x = 0$, we observe that the origin is a regular singular point of Bessel's equation, but is an ordinary point of Legendre's equation. Solution of Bessel's Equation that is given by reference Zill [5] is presented here with its coefficient c_{2n} where also is shown;

$$c_{2n} = \frac{(-1)^n}{2^{2n+v}n!(1+v)(2+v)\cdots(n+v)\Gamma(1+v)} = \frac{(-1)^n}{2^{2n+v}n!\Gamma(1+v+n)} \tag{C.27}$$

Using the coefficients c_{2n} and $r = v$, a series solution of Eq. (C.25) is $y = \sum_{n=0}^{\infty} c_{2n}x^{2n+v}$. This solution is usually denoted by $J_v(x)$:

$$J_v(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(1+v+n)}\left(\frac{x}{2}\right)^{2n+v} \tag{C.28}$$

If $v \geq 0$, the series converges at least on the interval $[0, \infty)$. Also, for the second exponent $r_2 = -v$ we obtain, in exactly the same manner,

$$J_{-v}(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(1-v+n)}\left(\frac{x}{2}\right)^{2n-v} \tag{C.29}$$

The function $J_v(x)$ and $J_{-v}(x)$ are called **Bessel Function of the First Kind** of order v and $-v$, respectively. Depending on the value of v, Equation C.29 may contain negative powers of and hence converge on $(0, \infty)$ [5]. When $v = 0$, it is apparent that Eqs. (C.28) and (C.29) are the same. If $v > 0$ and $r_1 - r_2 = v - (-v) = 2v$ is not a positive integer, it follows that $J_v(x)$ and $J_{-v}(x)$ are linearly independent solutions of Eq. (C.25) on $(0, \infty)$, and so the general solution of the interval is $y = c_1J_v(x) + c_2J_{-v}(x)$. But when $r_1 - r_2 = 2v$ is a positive integer, a second series solution of Eq. (C.25) may exist. In this case we distinguish two possibilities. When $v = m$ positive integer, $J_{-m}(x)$ defined by Eq. (C.29) and $J_m(x)$ are not linearly independent solution.

It can be shown that $J_{-m}(x)$ is a constant multiple of $J_m(x)$ [see (i) in properties of Bessel function within table below]. In addition, $r_1 - r_2 = 2v$ can a be a positive integer when $r_1 - r_2 = 2v$ is half and odd positive integer. In this case it can be

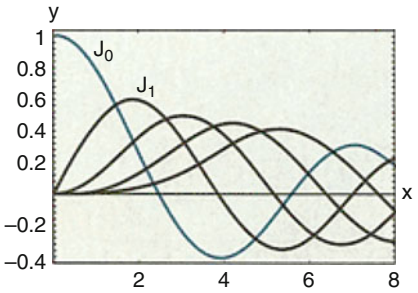


Fig. C.5 Bessel functions of the first kind for $n = 0,1,2,3,4$

shown that $J_v(x)$ and $J_{-v}(x)$ are linearly independent and the general solution of Eq. (C.25) on $(0,\infty)$ is given by

$$y = c_1 J_v(x) + c_2 J_{-v}(x) \quad v \neq \text{int } eger \tag{C.30}$$

The graph of $y = J_0(x)$ and $y = J_1(x)$ are given in Fig. C.5 below;

If v is not integer $(v \neq \text{int } eger)$, the function defined by the linear combination

$$Y_v(x) = \frac{\cos v\pi J_v(x) - J_{-v}(x)}{\sin v\pi} \tag{C.31}$$

and the function $J_v(x)$ are linearly independent solutions of Eq. (C.25). Thus another form of the general solution of Eq. (C.25) is $y = c_1 J_v(x) + c_2 Y_v(x)$, provided $v \neq \text{int } eger$. As $v \to m$, m an integer, Eq. (C.31) has the indeterminate form 0/0. However, it can be shown by L'Hospital's rule that $\lim_{v\to m} Y_v(x)$ exists. Moreover, the function

$$Y_m(x) = \lim_{v\to m} Y_v(x)$$

and $J_m(x)$ are linearly independent solution of $x^2y'' + xy' + (x^2 - m^2)y = 0$. Hence for *any value* of v the general solution of Eq. (C.25) on $(0,\infty)$ can be written as

$$y = c_1 J_v(x) + c_2 Y_v(x) \tag{C.32}$$

$Y_v(x)$ Is called the **Bessel Function of the Second Kind** of order v. Figure C.6 below shows the graph of $Y_0(x)$ and $Y_1(x)$.

By identifying $v^2 = 9$ and $v = 3$ we see from Eq. (C.32) that the general solution of the equation $x^2y'' + xy' + (x^2 - m^2)y = 0$ on $(0,\infty)$ is $y = c_1 J_3(x) + c_2 Y_3(x)$. Sometime it is possible to do some mathematical manipulation of a differential equation into form of Eq. (C.25) and be able to solve it as a Bessel Function approach as above [5].

Example 1: We have the following differential equation in form of $mx'' + ke^{-\alpha t}x = 0$, $\alpha > 0$. We can find the general solution of this equation using

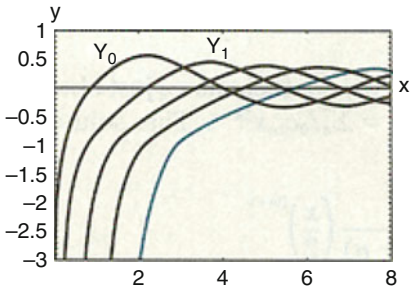


Fig. C.6 Bessel function of the second kind for $n = 0,1,2,3,4$

the Bessel Function Methods. This is the problem of free undamped motion of a mass on an aging spring.

Solution: If we change variables using the transformation of $s = \frac{2}{\alpha}\sqrt{\frac{k}{m}}$ then we have

$$s^2\frac{d^2x}{ds^2} + s\frac{dx}{ds} + s^2x = 0$$

This equation can be recognized if we set $v = 0$ in Eq. (C.25) and with the symbol x and s playing the roles of y and x, respectively. The general solution of the new equation is then given in form of $x = c_1J_0(s) + c_2Y_0(s)$. If we re-substitute s, the general solution of $mx'' + ke^{-\alpha t}x = 0$ is then seen to be [5].

$$x(t) = c_1J_0\left(\frac{2}{\alpha}\sqrt{\frac{k}{m}e^{-\alpha\frac{t}{2}}}\right) + c_2Y_0\left(\frac{2}{\alpha}\sqrt{\frac{k}{m}e^{-\alpha\frac{t}{2}}}\right)$$

Some Properties of Bessel Functions

We list below a few of the more useful properties of Bessel Functions of order m, m=01,1,2,. . ..

$$(i)\quad J_{-m}(x) = (-1)^mJ_m(x) \qquad (ii)\quad J_m(-x) = (-1)^mJ_m(x)$$

$$(iii)\qquad J_0(x) = \begin{cases} 0 & m > 0 \\ 1 & m = 0 \end{cases} \qquad (iv)\qquad \lim_{x\to 0} Y_m(x) = -\infty$$

Note that property (ii) indicates that $J_m(x)$ is an even function if m is an even integer and odd function if m is an odd integer. The graphs of $Y_0(x)$ and $Y_1(x)$

(continued)

in Fig. C.6 illustrate property (*iv*): is unbounded at the origin. This last fact is not obvious from Eq. (C.31). It can be shown either from Eq. (C.31) or by the methods of solutions about singular point 5 that for,

$$Y_0(x) = \frac{2}{\pi} J_0(x) \left[\gamma + \ln \frac{x}{2} \right] - \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^k}{(k!)^2} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k} \right) \left(\frac{x}{2} \right)^{2k}$$

where $\gamma = 0.57721566\ldots$. is **Euler's Constant**. Because of the presence of the logarithmic term, $Y_0(x)$ is discontinuous at $x = 0$.

C.2 Initial-Value Problem (IVP)

We are often interested in solving a differential equation subject to prescribed side conditions. Conditions that are imposed on the unknown solution $y = y(x)$ or its derivatives. On some interval I containing x_0, the problem

$$\text{solve}: \quad \frac{d^n y}{dx^n} = f\left(x, y, y', \cdots, y^{(n-1)} \right) \tag{C.33}$$

$$\textit{Subject to}: \quad y(x_0) = y_0 \quad y'(x_0) = y_1, \cdots, y^{(n-1)}(x_0) = y_{n-1} \tag{C.34}$$

where $y_1, y_2, \cdots \bullet, y_{n-1}$ are arbitrarily specified real constant is called **Initial-Value Problem (IVP)**. The values of $y(x)$ and its first $n-1$ derivative at a single point $x_0 : y(x_0) = y_0,\ y'(x_0) = y_1, \cdots, y^{(n-1)}(x_0) = y_{n-1}$ are called **Initial Conditions**. The problem given in Eq. (C.33) in above also called an ***n*th-order initial-value problem**. For example

$$\begin{aligned} &\text{Solve}: \quad \frac{dy}{dx} = f(x, y) \\ &\text{Subject}: \quad y(x_0) = y_0 \end{aligned} \tag{C.35}$$

$$\begin{aligned} &\text{Solve}: \quad \frac{d^2 y}{dx^2} = f\left(x, y, y' \right) \\ &\text{Subject}: \quad y(x_0) = y_0 \quad \text{and} \quad y'(x_0) = y_1 \end{aligned} \tag{C.36}$$

are first and second-ordered initial-value problem, respectively. These two problems are easy to interpret in geometric terms. For Eq. (C.35) we are seeking a solution of the differential equation on an interval I containing x_0 so that a solution curve passes through the prescribed point (x_0, y_0), see Fig. C.7.

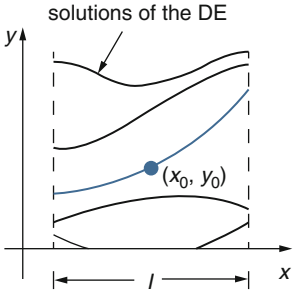


Fig. C.7 First-Order IVP

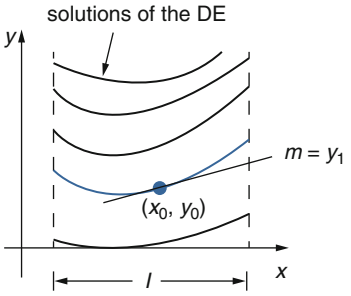


Fig. C.8 Second-Order IVP

For Eq. (C.36) we want to find a solution of the differential equation whose graph not only passes through (x_0, y_0) but passes through so that the slope of the curve at this point is y_1. See Fig. C.8. The term initial condition derives from physical system where the independent variable is time t and where $y(t_0) = y_0$ and $y'(t_0) = y_1$ represent, respectively [5], the position and velocity of an object at some beginning, or initial, time t_0. Solving an nth-order initial-value problem frequently entails using an n-parameter family of solutions of the given differential equation to find n specialized constants so that the resulting par solution of the equation also "fits" that is, satisfies the n initial conditions [5].

Example 1: First-Order IVP It can be verified that $y = ce^x$ is a one-parameter family of solutions of the simple first-order equation $y' = y$ on the interval $(-\infty, \infty)$. If we assume an initial condition such as $y(0) = 3$, then substitute $x = 0$, $y = 3$ in the family determines the constant $3 = ce^0 = c$. Thus the function $y = 3e^x$ is a solution of the initial-value problem.

We can pick another sets of value for x and y say $y(1) = -2$ to obtain another set of solution for the family such as $y = -2e^{x-1}$ and Fig. C.9 can be produced.

Example 2: Second-Order IVP If we pick another example such as $x = c_1 \cos 4t + c_2 \sin 4t$, we can observe a two-values of IPV is need in order to solve this family. If we let $x\left(\frac{\pi}{2}\right) = -2$ and $x'\left(\frac{\pi}{2}\right) = 1$ then with a little mathematics work and taking first derivative of the example we will find the following solution and values for c_1 and c_2 as follows:

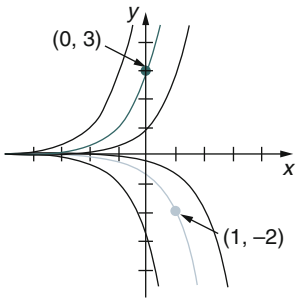


Fig. C.9 Solution of IVPs family

$$\begin{cases} c_1 = -2 \quad \text{for} \quad t = \dfrac{\pi}{2} \\ \\ c_2 = \dfrac{1}{4} \quad \text{for} \quad t = -\dfrac{\pi}{2} \end{cases} \quad \text{then the solution is} \quad x = -2 \cos 4t + \frac{1}{4}\sin 4t$$

Summarizing the IVP problem we have to define the *Intervals of Validity*. First let us take a look at a theorem about linear first-order differential equations. This is a very important theorem although it is not going to be really used for its most important aspect.

Theorem 1
Consider the following IVP.

$$y' + p(t)y = g(t) \quad y(t_0) = y_0$$

If $p(t)$ and $g(t)$ are continuous functions on an open interval $\alpha < t < \beta$ and the interval contains t_0, then there is a unique solution to the IVP on that interval.

This theorem tells us, first for a linear first-order differential equations a solutions are definitely exists and more importantly the solution is a unique one. We may not be able to find the solution, but we do know that it exists and there will only be one of them. This is a very important aspect of this theorem.

Next, if the interval in the theorem is the largest possible interval on which $p(t)$ and $g(t)$ are continuous, then the interval is the interval of validity for the solution. This means, that for linear first-order differential equations, we will not need to actually solve the differential equation in order to find the interval of validity. Notice as well that the interval of validity will dependent only partially on the initial condition. The interval must contain t_0, but the value of y_0, has no effect on the interval of validity.

Example 1: Without solving, determine the interval of validity for the following initial value problem.

$$(t^2 - 9)y' + 2y = \ln|20 - 4t| \quad y(4) = -3$$

Solution: First, in order to use the theorem to find the interval of validity we must write the differential equation in the proper form given in the theorem. So we will need to divide out by the coefficient of the derivative.

$$y' + \frac{2}{t^2 - 9}y = \frac{\ln|20 - 4t|}{t^2 - 9}$$

Next, we need to identify where the two functions are not continuous. This will allow us to find all possible intervals of validity for the differential equation. So, $p(t)$ will be discontinuous at $t = \pm 3$ since these points will give a division by zero. Likewise, $g(t)$ will also be discontinuous at $t = \pm 3$ as well as $t = 5$ since at this point we will have the natural logarithm of zero. Note that in this case we won't have to worry about natural log of negative numbers because of the absolute values.

Now, with these points in hand we can break up the real number line into four intervals where both $p(t)$ and $g(t)$ will be continuous. These four intervals are

$$-\infty < t < -3 \qquad -3 < t < 3 \qquad 3 < t < 5 \qquad 5 < t < \infty$$

The endpoints of each of the intervals are points where at least one of the two functions is discontinuous. This will guarantee that both functions are continuous everywhere in each interval. Finally, let's identify the actual interval of validity for the initial value problem. The actual interval of validity is the interval that will contain $t_0 = 4$. So, the interval of validity for the initial value problem is.

$$3 < t < 5$$

In this last example we need to be careful to not jump to the conclusion that the other three intervals cannot be intervals of validity. By changing the initial condition, in particular the value of to, we can make any of the four intervals the interval of validity. The first theorem required a linear differential equation. There is a similar theorem for non-linear first-order differential equations. This theorem is not as useful for finding intervals of validity as the first theorem was so we won't be doing all that much with it.

Here is the theorem.

Theorem 2
Consider the following IVP

$$y' = f(t,y) \quad y(t_0) = y_0$$

If $f(t,y)$ and $\frac{\partial f}{\partial y}$ are continuous function in some rectangle $\alpha < t < \beta$, $\gamma < y < \delta$ containing the point (t_0, y_0) then there is a unique solution to the IVP in some interval $t_0 - h < t < t_0 + h$ that is contained in $\alpha < t < \beta$.

Unlike the first theorem, this one cannot really be used to find an interval of validity. So, we will know that a unique solution exists if the conditions of the theorem are met, but we will actually need the solution in order to determine its interval of validity. Note as well that for non-linear differential equations it appears that the value of y_0 may affect the interval of validity. Here is an example of the problems that can arise when the conditions of this theorem aren't met.

Example 2: Determine all possible solutions to the following IVP.

$$y' = y^{\frac{1}{5}} \quad y(0) = 0$$

Solution: First, notice that this differential equation does NOT satisfy the conditions of the theorem.

$$f(y) = y^{\frac{1}{5}} \quad \frac{df}{dy} = \frac{1}{3y^{\frac{2}{5}}}$$

So, the function is continuous on any interval, but the derivative is not continuous at $y = 0$ and so will not be continuous at any interval containing $y = 0$. In order to use the theorem both must be continuous on an interval that contains $y_0 = 0$ and this is problem for us since we do have $y_0 = 0$. This differential equation is separable and is fairly simple to solve.

$$\int y^{-\frac{1}{5}} dy = \int dt$$

$$\frac{3}{2} y^{\frac{2}{3}} = t + c$$

Applying the initial condition gives $c = 0$ and so the solution is.

$$\frac{3}{2}y^{\frac{2}{3}} = t$$

$$y^{\frac{2}{3}} = \frac{2}{3}t$$

$$y^2 = \left(\frac{2}{3}t\right)^3$$

$$y(t) = \pm\left(\frac{2}{3}t\right)^{\frac{3}{2}}$$

We have got two possible solutions here, both of which satisfy the differential equation and the initial condition. There is also a third solution to the IVP. $y(t) = 0$ is also a solution to the differential equation and satisfies the initial condition.

In this last example we had a very simple IVP and it only violated one of the conditions of the theorem, yet it had three different solutions. All the examples we've worked in the previous sections satisfied the conditions of this theorem and had a single unique solution to the IVP. This example is a useful reminder of the fact that, in the field of differential equations, things don't always behave nicely. It's easy to forget this as most of the problems that are worked in a differential equations class are nice and behave in a nice, predictable manner.

Let's work one final example that will illustrate one of the differences between linear and non-linear differential equations.

Example 3: Determine the interval of validity for the initial value problem below and give its dependence on the value of y_0.

$$y' = y^2 \quad y(0) = y_0$$

Solution: Before proceeding in this problem, we should note that the differential equation is non-linear and meets both conditions of the Theorem 2 and so there will be a unique solution to the IVP for each possible value of y_0.

Also, note that the problem asks for any dependence of the interval of validity on the value of y_0. This immediately illustrates a difference between linear and non-linear differential equations. Intervals of validity for linear differential equations do not depend on the value of y_0. Intervals of validity for non-linear differential can depend on the value of y_0 as we pointed out after the second theorem.

So, let's solve the IVP and get some intervals of validity.

First note that if $y_0 = 0$ then $y(t) = 0$ is the solution and this has an interval of validity of $-\infty < t < \infty$.

So for the rest of the problem let's assume that $y_0 \neq 0$. Now, the differential equation is separable so let's solve it and get a general solution.

$$\int y^{-2} dy = \int dt$$

$$-\frac{1}{y} = t + c$$

Applying the initial condition gives

$$c = -\frac{1}{y_0}$$

The solution is then.

$$-\frac{1}{y} = t + -\frac{1}{y_0}$$

Now that we have a solution to the initial value problem we can start finding intervals of validity. From the solution we can see that the only problem that we'll have is division by zero at

This leads to two possible intervals of validity.

$$-\infty < t < \frac{1}{y_0}$$

$$\frac{1}{y_0} < t < \infty$$

That actual interval of validity will be the interval that contains $t_0 = 0$. This however, depends on the value of y_0. If $y_0 < 0$ then $\frac{1}{y_0} < 0$ and so the second interval will contain $t_0 = 0$. Likewise if $y_0 > 0$ then $\frac{1}{y_0} > 0$ and in this case the first interval will contain $t_0 = 0$.

This leads to the following possible intervals of validity, depending on the value of yo.

If $y_0 > 0$ $\quad -\infty < t < \frac{1}{y_0}$ is the interval of validity.

If $y_0 = 0$ $\quad -\infty < t < \infty$ is the interval of validity.

If $y_0 < 0$ $\quad \frac{1}{y_0} < t < \infty$ is the interval of validity.

On a side note, notice that the solution, in its final form, will also work if $y_0 = 0$.

So what did this example show us about the difference between linear and non-linear differential equations?

First, as pointed out in the solution to the example, intervals of validity for non-linear differential equations can depend on the value of y_0, whereas intervals of validity for linear differential equations don't.

Second, intervals of validity for linear differential equations can be found from the differential equation with no knowledge of the solution. This is definitely not the case with nonlinear differential equations. It would be very difficult to see how any of these intervals in the last example could be found from the differential equation. Knowledge of the solution was required in order for us to find the interval of validity.

C.3 Boundary-Value Problems (BVP)

Before we start off this section we need to make it very clear that we are only going to scratch the surface of the topic of boundary value problems. There is enough material in the topic of boundary value problems that we could devote a whole class to it. The intent of this section is to give a brief (and we mean very brief) look at the idea of boundary value problems and to give enough information to allow us to do some basic partial differential equations in the next chapter.

Now, with that out of the way, the first thing that we need to do is to define just what we mean by a boundary value problem (BVP for short). With initial value problems we had a differential equation and we specified the value of the solution and an appropriate number of derivatives at the same point (collectively called initial conditions). For instance for a second-order differential equation the initial conditions are

$$y(t_0) = y_0 \quad y'(t_0) = y_0'$$

With boundary value problems we will have a differential equation and we will specify the function and/or derivatives at different points, which we'll call boundary values. For second order differential equations, which will be looking at pretty much exclusively here, any of the following can, and will, be used for boundary conditions.

$$y(x_0) = y_0 \quad y(x_1) = y_1 \tag{C.37}$$

$$y'(x_0) = y_0 \quad y'(x_1) = y_1 \tag{C.38}$$

$$y'(x_0) = y_0 \quad y(x_1) = y_1 \tag{C.39}$$

$$y(x_0) = y_0 \quad y'(x_1) = y_1 \tag{C.40}$$

As mentioned above we'll be looking pretty much exclusively at second order differential equations. We will also be restricting ourselves down to linear differential equations. So, for the purposes of our discussion here we'll be looking almost exclusively at differential equations in the form,

$$y'' + p(x)y' + q(x)y = g(x) \tag{C.41}$$

along with one of the sets of boundary conditions given in Eqs. (C.37–C.40). We will, on occasion, look at some different boundary conditions but the differential equation will always be on that can be written in this form.

As we'll soon see much of what we know about initial value problems will not hold here. We can, of course, solve Eq. (C.41) provided the coefficients are constants and for a few cases in which they aren't. None of that will change. The changes (and perhaps the problems) arise when we move from initial conditions to boundary conditions.

One of the first changes is a definition that we saw all the time in the earlier chapters. In the earlier chapters we said that a differential equation was homogeneous if $g(x) = 0$ for all x. Here we will say that a boundary value problem is homogeneous if in addition to $g(x) = 0$ we also have $y_0 = 0$ and $y_1 = 0$ (regardless of the boundary conditions we use). If any of these are not zero we will call the BVP nonhomogeneous.

It is important to now remember that when we say homogeneous (or nonhomogeneous) we are saying something not only about the differential equation itself but also about the boundary conditions as well.

The biggest change that we're going to see here comes when we go to solve the boundary value problem. When solving linear initial value problems a unique solution will be guaranteed under very mild conditions. We only looked at this idea for first-order IVP's but the idea does extend to higher order IVP's. In that section we saw that all we needed to guarantee a unique solution was some basic continuity conditions. With boundary value problems we will often have no solution or infinitely many solutions even for very nice differential equations that would yield a unique solution if we had initial conditions instead of boundary conditions.

Before we get into solving some of these let's next address the question of why we're even talking about these in the first place. As we'll see in the next chapter in the process of solving some partial differential equations we will run into boundary value problems that will need to be solved as well. In fact, a large part of the solution process there will be in dealing with the solution to the BVP. In these cases the boundary conditions will represent things like the temperature at either end of a bar, or the heat flow into/out of either end of a bar. Or maybe they will represent the location of ends of a vibrating string. So, the boundary conditions there will really be conditions on the boundary of some process.

So, with some of basic stuff out of the way let's find some solutions to a few boundary value problems. Note as well that there really isn't anything new here yet. We know how to solve the differential equation and we know how to find the

constants by applying the conditions. The only difference is that here we'll be applying boundary conditions instead of initial conditions.

Example 1: Solve the following BVP. $y'' + 4y = 0$ with $y(0) = -2$ and $y\left(\frac{\pi}{4}\right) = 10$.

Solution: This a simple differential equation to solve and the solution is in the form of the following as a general solution. Finding these types of solutions is explained in the next section.

$$y(x) = c_1 \cos(2x) + c_2 \sin(2x)$$

Now all that we need to do is apply the boundary conditions.

$$-2 = y(0) = c_1$$

$$10 = y\left(\frac{\pi}{4}\right) = c_2$$

Then the solution is $y(x) = -2\cos(2x) + 10\sin(2x)$

We mentioned above that some boundary value problems can have no solutions or infinite solutions we had better do a couple of examples of those as well here. This next set of examples will also show just how small of a change to the BVP it takes to move into these other possibilities.

Example 2: Solve the BVP of $y'' + 4y = 0$ and with initial condition of $y(0) = -2$ and $y(2\pi) = -2$.

Solution: We are working with the same differential equation as the first example so we still have,

$$y(x) = c_1 \cos(2x) + c_2 \sin(2x)$$

Upon applying the boundary conditions we get,

$$-2 = y(0) = c_1$$

$$-2 = y(2\pi) = c_1$$

So in this case, unlike the previous example, both boundary conditions tell us that have to have and neither one of them tell us anything about . Remember however that all we are asking for is a solution to the differential equation that satisfies the two given boundary conditions and the following function will do that,

$$y(x) = -2\cos(2x) + c_2 \sin(2x)$$

In other words, regardless of the value of c_2 we get a solution and so, in this case we infinitely solutions to the boundary vale problem. For information and example refer to the following web site.

http://tutorial.math.lamar.edu/Classes/DE/BoundaryValueProblem.aspx

C.3.1 Second-Order Linear Equation

Linear equations of second order are of crucial importance in the study of differential equations for two main reasons. The first is that linear equations have a rich theoretical structure that underlies a number of systematic methods of solution. Further, a substantial portion of this structure and these methods are understandable at a fairly elementary mathematical level. In order to present the key ideas in the simplest possible context, we describe them in this chapter for second-order equations. Another reason to study second order linear equations is that they are vital to any serious investigation of the classical areas of mathematical physics. One cannot go very far in the development of fluid mechanics, heat conduction, wave motion, or electromagnetic phenomena without finding it necessary to solve second-order linear differential equations.

C.3.2 Homogeneous Linear Equations with Constant Coefficients

We have seen that the first-order differential equation of the form $y^{'} + ay = 0$ where a is a constant, possesses the exponential solution $y = c_1 e^{-ax}$ on the interval $(-\infty, \infty)$. Therefore, it is natural to ask whether exponential solution exists on $(-\infty, \infty)$ for homogeneous linear higher-order differential equations such as the following:

$$a_n y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_2 y^{''} + a_1 y^{'} + a_0 y = 0 \tag{C.42}$$

Where the coefficients $a, i = 0, 1, \cdots, n$ are real constant and $a_n \neq 0$. The surprising fact is that all solutions of Eq. (C.42) are exponential function or are constructed out of exponential [5] functions.

A second-order ordinary differential equation as we said in general with both homogeneous and non-homogeneous parts is presented below as Eq. (C.43).

$$\frac{d^2 y}{dt^2} = f\left(t, y, \frac{dy}{dt}\right) \tag{C.43}$$

Where f is some given function and can be considered the non-homogeneous part of the second-order differential equation while $\frac{d^2y}{dt^2} = 0$ is the special case or auxiliary form of second order known as homogeneous part of Eq. (C.43). This equation has a general solution such as (y) that consists of solution of both terms and they can be designated (y_c) which is Auxiliary or Complementary solution its homogeneous part while (y_p) know as Particular solution for non-homogeneous part of same equation. This solution is shown as follows:

$$y = y_c + y_p \tag{C.44}$$

Usually, we will denote the independent variable by t since time is often the independent variable in physical problems, but sometimes we will use x instead, while letter of y is used as dependent variable. Eq. (C.43) is linear ordinary differential equation if f has the form of following Eq. (C.45), that is, if f is linear in y and y'.

$$\begin{aligned} f\left(t, y, \frac{dy}{dt}\right) &= g(t) - p(t)\frac{dy}{dt} - q(t)y \\ &= g(t) - p(t)y' - q(t)y \end{aligned} \tag{C.45}$$

In Equation C.45 g, p, and q are specific functions of the independent variable but do not dependent on . In this we can rewrite Eq. (C.45) as

$$y'' + p(t)y' + q(t)y = g(t) \tag{C.46}$$

Sometime Eq. (C.46) can be written as Eq. (C.47) below and if $P(t) \neq 0$, then we can divide both side of this equation by $P(t)$ and obtain Eq. (C.46).

$$P(t)y'' + Q(t)y' + R(t)y = G(t) \tag{C.47}$$

Where in this case we have,

$$p(t) = \frac{Q(t)}{P(t)} \quad q(t) = \frac{R(t)}{P(t)} \quad g(t) = \frac{G(t)}{P(t)} \tag{C.48}$$

If Equation C.43 is not of the form Eqs. (C.46) or (C.47), then it is called **nonlinear**. Finding an analytical solution of nonlinear equations are not trivial and requires numerical or geometrical approaches that are more often appropriate. Additionally there two more special types of second-order nonlinear equations that can be solved by a change of variables that reduces them to first-order equations. In most cases an initial value problem (IVP) consisting type of differential equations such as Eqs. (C.43), (C.46), and (C.47) involves a set of initial conditions such as $y(t_0) = y_0$ and $y'(t_0) = y_0'$, where y_0 and y' are given numbers. The initial conditions for a

second-order equation prescribe not only a particular point (t_0, y_0) through which the graph of the solution must pass, but also the slope $y_0^{'}$ of the graph at that point. It is reasonable to expect that two initial conditions are needed for a second-order equation because, roughly speaking, two integrations are required to find a solution and each integration introduces an arbitrary constant. Presumably, two initial conditions will suffice to determine values for these two constants.

A second-order linear equation is said to be homogeneous if the term $g(t)$ in Eq. (C.46), or the term $G(t)$ in Eq. (C.47), is zero for all t. Otherwise, the equation is called non-homogeneous. As a result, the term $g(t)$, or $G(t)$, is sometimes called the non-homogeneous term. We begin our discussion with homogeneous equations, which we will write in the form

$$P(t)y^{''} + Q(t)y^{'} + R(t)y = 0 \tag{C.49}$$

First we concentrate our attention to special case where the P, Q, and R are constants and in that case Eq. (C.49) becomes in the form of Eq. C.50 and we call it **Auxiliary Equation**, which is a very special case of Equation of (C.43) and as a result Eq. (C.42).

$$ay^{''} + by^{'} + cy = 0 \tag{C.50}$$

If we try a solution of the form $y = e^{mt}$, then after substituting $y^{'} = me^{mt}$ and $y^{''} = m^2e^{mt}$ into Eq. (C.50), this equation takes the following form:

$$am^2e^{mt} + bme^{mt} + ce^{mt} = 0$$

or

$$e^{mt}\left(am^2 + bm + c\right) = 0$$

Since e^{mt} is never zero for real values of t, it is apparent that the only way this exponential function can satisfy the differential Eq. (C.50) is if m is chosen as a root of the quadratic equation

$$am^2 + bm + c = 0 \tag{C.51}$$

This last equation is called the **auxiliary or characteristic equation** of the differential Eq. (C.50) and has two roots as follows:

$$m_1 = \frac{\left(-b + \sqrt{b^2 - 4ac}\right)}{2a}$$

$$m_2 = \frac{\left(-b - \sqrt{b^2 - 4ac}\right)}{2a}$$

There will be three forms of the general solution of Eq. (C.50) corresponding to the three cases;

1. m_1 and m_2 real and positive $\left(b^2 - 4ac\right) > 0$.
2. m_1 and m_2 real and equal $\left(b^2 - 4ac\right) = 0$.
3. m_1 and m_2 conjugate complex numbers $\left(b^2 - 4ac\right) < 0$.

Case 1: Distinct Real Roots. Under the assumption that the auxiliary Eq. (Eq. C.51) has two unequal real roots m_1 and m_2, we find two solutions, $y_1(t) = e^{m_1 t}$ and $y_2(t) = e^{m_2 t}$. These functions are linearly independent on $(-\infty, +\infty)$ and hence form a fundamental set. It follows that the general solution of Eq. (C.50) on this interval is

$$y = c_1 y_1(t) + c_2 y_2(t) = c_1 e^{m_1 t} + c_2 e^{m_2 t} \tag{C.52}$$

Example 1: Solve equation $y'' - y = 0$ that satisfies the initial conditions of $y(0) = 2$ and $y'(0) = -1$.

Solution: Comparing the differential equation of the example with Eq. (C.50) we see that $a = 1$, $b = 0$, and $c = -1$. So we looking for a solution that not only satisfies the example problem also pass through the point (0, 2) and at that point has the slope -1. The first solution is the type two distinct real roots of $m_1 = 1$ and $m_2 = -1$. As a result we have two separate solutions of $y_1(t) = c_1 e^t$ and $y_2(t) = c_2 e^{-t}$ with general solution of $y = c_1 y_1(t) + c_2 y_2(t) = c_1 e^t + c_2 e^{-t}$ form. Secondly, we set $t = 0$ and $y = 2$ in this general solution and we get $c_1 + c_2 = 2$. Next, we differentiate the general solution and we have $y' = c_1 e^t - c_2 e^{-t}$. Then setting $t = 0$ and $y' = 1$, we obtain $c_1 - c_2 = -1$. By solving for c_1 and c_2, we get that $c_1 = \frac{1}{2}$ and $c_2 = \frac{3}{2}$. Finally inserting these results in general solution we have the final solution of the example which is $y = \frac{1}{2} e^t + c_2 e^{-t}$.

Example 2: Solve equation $y'' + 5y' + 6y = 0$ that satisfies the initial conditions of $y(0) = 2$ and $y'(0) = 3$.

Solution: We are not going to explain all the steps but show all the details. Assume $y = e^{rt}$ then we have;

$$r^2 + 5r + 6 = (r+2)(r+3) = 0$$

$$r_1 = -2 \quad \text{and} \quad r_2 = -3$$

$$y = c_1 e^{-2t} + c_2 e^{-3t}$$

Using the initial condition we have

$$\begin{cases} c_1 + c_2 = 2 \\ \\ -2c_1 - 3c_2 = 3 \end{cases}$$

$$\begin{cases} c_1 = 9 \\ \\ c_2 = -7 \end{cases}$$

$$y = 9e^{-2t} - 7e^{-3t}$$

The graph of the solution is shown in Fig. C.10.

Case 2: Repeated Real Roots. When $m_1 = m_2$ we necessarily obtain only one exponential solution, $y_1(t) = e^{m_1 t}$. From the quadratic formula we find that $m_1 = -b/2a$ since the only way to have $m_1 = m_2$ is to have $\left(b^2 - 4ac\right) = 0$. It follows that a second solution of the equation is [5];

$$y_2 = e^{m_1 t} \int \frac{e^{2m_1 t}}{e^{2m_1 t}} dt = e^{m_1 t} \int dt = t e^{m_1 t} \tag{C.53}$$

In Eq. (C.53) we have used the fact that $-b/a = 2m_1$. The general solution is then

$$y = c_1 e^{m_1 t} + c_2 t e^{m_1 t} \tag{C.54}$$

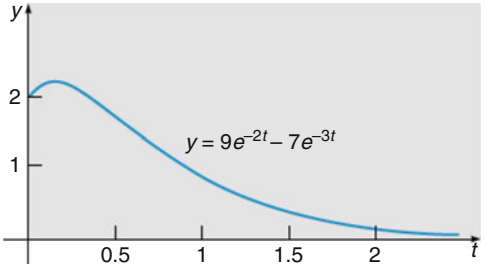


Fig. C.10 Solution of $y'' + 5y' + 6y = 0$, $y(0) = 2, y'(0) = 3$

Example 1 Solve the differential equation $y'' - 10y' + 25y = 0$.

Solution: Again assume $y = e^{mt}$ then we have

$$m^2 - 10m + 25 = (m-5)^2 = 0$$

$$m_1 = m_2 = 5$$

Then

$$y = c_1 e^{5t} + c_2 t e^{5t}$$

Case 3: Complex Roots. If m_1 and m_2 are complex, then we can write $m_1 = \alpha + i\beta$ and $m_2 = \alpha - i\beta$, where α and $\beta > 0$ are real and $i^2 = -1$. Formally, there is no difference between this case and Case 1, and hence

$$y = c_1 e^{(\alpha+i\beta)t} + c_2 e^{(\alpha-i\beta)t}$$

Using Euler's formula we have;

$$e^{i\theta} = \cos\ \theta + i\ \sin\ \theta$$

where θ is any real number. It follows from this formula that

$$e^{i\beta t} = \cos\ \beta t + i\ \sin\ \beta t \quad \text{and} \quad e^{-i\beta t} = \cos\ \beta t - i\ \sin\ \beta t \tag{C.55}$$

where we have used $\cos(-\beta t) = \cos\ \beta t$ and $\sin(-\beta t) = -\sin\ \beta t$. By adding first and then subtracting the two equations in Eq. (C.55), we obtain, respectively,

$$e^{i\beta t} + e^{-i\beta t} = 2\cos\ \beta t \quad \text{and} \quad e^{i\beta t} - e^{-i\beta t} = 2\ i\sin\ \beta t$$

Since $y = c_1 e^{(\alpha+i\beta)t} + c_2 e^{(\alpha-i\beta)t}$ is a solution of Eq. (C.50) for any choice of the constant c_1 and c_2, the choices $c_1 = c_2 = 1$ and $c_1 = 1$, $c_2 = -1$ give, in turn, two solutions [4]:

$$y_1 = c_1 e^{(\alpha+i\beta)t} + c_2 e^{(\alpha-i\beta)t} \quad \text{and} \quad y_2 = c_1 e^{(\alpha+i\beta)t} - c_2 e^{(\alpha-i\beta)t}$$

But $y_1 = e^{\alpha t}\left(e^{i\beta t} + e^{-i\beta t}\right) = 2e^{\alpha t}\ \cos\ \beta t$

and $y_2 = e^{\alpha t}\left(e^{i\beta t} - e^{-i\beta t}\right) = 2ie^{\alpha t}\ \sin\ \beta t$

The last two results show that $e^{\alpha t}\ \cos\ \beta t$ and $e^{\alpha t}\ \sin\ \beta t$ are *real* solutions of Eq. (C.50). Moreover, these solutions form a fundamental set on $(-\infty, \infty)$. As a result, the general solution is

$$y = c_1 e^{\alpha t}\cos\beta t + c_2 e^{\alpha t}\sin\beta t = e^{\alpha t}(c_1\ \cos\ \beta t + c_2\ \sin\ \beta t) \tag{C.56}$$

Example 1: Solve the differential equation $y'' + 4y' + 7y = 0$.

Solution: Again assume $y = e^{mt}$ then we have

$$m^2 + 4m + 7 = 0$$

$$m_1 = -2 + \sqrt{3i} \quad \text{and} \quad m_2 = -2 - \sqrt{3i}$$

Then with $\alpha = -2$ and $\beta = \sqrt{3}$, we have

$$y = e^{-2t}\left(c_1 \cos \sqrt{3t} + c_2 \sin \sqrt{3t}\right)$$

C.3.3 *Fundamental Solutions of Linear Homogeneous Equations*

In the preceding section it was shown how to solve some differential equations of the form Eq. (C.50) as follows:

$$ay'' + by' + cy = 0$$

where a, b, and c are constants. Now we build on those results to provide a clearer picture of the structure of the solutions of all second order linear homogeneous equations.

Theorem 1
Consider the initial value problem

$$y'' + p(t)y' + q(t)y = g(t) \quad y(t_0) = y_0 \quad \text{and} \quad y'(t_0) = y_0' \tag{C.57}$$

where p, q, and g are continuous on an open interval I. Then there is exactly one

solution $y = \phi(t)$ of this problem, and the solution exist throughout the interval I. This emphasize three thing as follows:

1. The initial value problem has a solution; in other words, a solution *exists*.
2. The initial value problem has only one solution; that is, the solution is *unique*.
3. The solution ϕ is defined *throughout the interval I* where the coefficients are continuous and is at least twice differentiable there.

For some problem of these assertion we have seen solution in example previous section such as $y'' - y = 0$ and initial value of $y(0) = 2$ and $y'(0) = -1$ which has solution of $y = \frac{1}{2}e^{t} + \frac{3}{2}e^{-t}$. The above *Theorem 1* indicates that this solution is indeed the only solution of the initial value problem (IVP) [5]. However, for most problems of the form Eq. (C.57), it is not possible to write down a useful expression for the solution. This is a major difference between first order and second order linear equations. For more information please refer to reference [6] of this appendix where you can find some unique techniques such as **Wronskian Determinant**, or simply **Wronskian**. Here we show a brief version of **Principle of Superposition** along with few examples which is in the form of the following equation [5].

$$y'' + p(t)y' + q(t)y = 0 \quad y(t_0) = y_0 \quad y'(t_0) = y_0' \tag{C.58}$$

Example 1: Find the solution of Eq. (C.58) in above where p and q are continuous in an open interval I containing t_0.

Solution: The solution of this problem is given by Boyce and DiPrima of reference [5] in this appendix. The summary of this solution is presented here

$$y = c_1 y_1(t) + c_2 y_2(t) \tag{C.59}$$

Where c_1 and c_2 are give as follows;

$$c_1 = \frac{y_0 y_2'(t_0) - y_0' y_2(t_0)}{y_1(t_0)y_2'(t_0) - y_1'(t_0)y_2(t_0)} \quad \text{and} \quad c_2 = \frac{-y_0 y_1'(t_0) - y_0' y_1(t_0)}{y_1(t_0)y_2'(t_0) - y_1'(t_0)y_2(t_0)}$$

$$c_1 = \frac{\begin{vmatrix} y_0 & y_2(t_0) \\ y_0' & y_2'(t_0) \end{vmatrix}}{\begin{vmatrix} y_1(t_0) & y_2(t_0) \\ y_1'(t_0) & y_2'(t_0) \end{vmatrix}} \quad \text{and} \quad c_2 = \frac{\begin{vmatrix} y_1(t_0) & y_0 \\ y_0'(t_0) & y_2' \end{vmatrix}}{\begin{vmatrix} y_1(t_0) & y_2(t_0) \\ y_1'(t_0) & y_2'(t_0) \end{vmatrix}}$$

With these values for c_1 and c_2 the Eq. (C.58) satisfies its initial condition as well as the differential equation in it.

Example 2: Given differential equation of $y'' + 5y' + 6y = 0$ find the Wronskian of y_1 and y_2.

Solution: The solution for this differential equation is given by $y_1(t) = e^{-2t}$ and $y_2(t) = e^{-3t}$ where readers can work out the details of these two answers by themselves. Then Wronskian of these two functions is

$$W = \begin{vmatrix} e^{-2t} & e^{-3t} \\ -2e^{-2t} & -3e^{-3t} \end{vmatrix} = -e^{-5t}$$

C.3.4 *Nonhomogeneous Equations; Method of Undetermined Coefficients*

Now if we return to Eq. (C.46) and assume a Nonhomogeneous aspect of this equation we then have

$$L(y) = y'' + p(t)y' + q(t)y = g(t) \tag{C.60}$$

Where p, q, and g are continuous functions on the open interval I. The equation

$$L(y) = y'' + p(t)y' + q(t)y = 0 \tag{C.61}$$

in which $g(t) = 0$ and p and q are the same as in Eq. (C.60), is called the homogeneous equation corresponding to Eq. (C.60). The following two results describe the structure of solutions of the nonhomogeneous Eq. (C.60) and provide a basis for constructing its general solution [5].

Theorem I
If Y_1 and Y_2 are two solutions of the nonhomogeneous Eq. (C.60), then their difference $Y_1 - Y_2$ is a solution of the corresponding homogeneous Eq. (C.61). If, in addition y_1, and y_2 are a fundamental set of solutions of Eq. (C.61), then

$$Y_1(t) - Y_2(t) = c_1 y_1(t) + c_2 y_2(t) \tag{C.62}$$

where c_1 and c_2 are certain constants [6].

The proof of this theorem is given by Boyce and DiPrima of reference [5] in this appendix.

Theorem II
The general solution of the nonhomogeneous Eq. (C.60) can be written in form

$$y = \phi(t) = c_1 y_1(t) + c_2 y_2(t) + Y(t) \tag{C.63}$$

where y_1 and y_2 are fundamental set of solutions of the corresponding homogeneous equation of Eq. (C.61), c_1 and c_2 are arbitrary constants, and Y is some specific solution of the nonhomogeneous Eq. (C.60) [6].

The proof of this theorem is give by Boyce and DiPrima of reference [6] in this appendix. The ***Theorem II*** states that to solve the nonhomogeneous Eq. (C.60), we must do three steps:

1. Find the general solution $c_1 y_1(t) + c_2 y_2(t)$ of the corresponding homogeneous equation. This solution is frequently called the ***Complementary Solution*** and may be denoted by $y_c(t)$.
2. Find some single solution $Y(t)$ of the nonhomogeneous equation. Often this solution is referred to as a ***Particular Solution*** and denoted by $y_p(t)$.
3. Add the two solutions in the two preceding steps we have a solution for Eq. (C.61) that denoted as $y(t) = y_c(t) + y_p(t)$.

We have presented how to find a $y_c(t)$ solution, at least when the homogeneous Eq. (C.61) has constant coefficients in previous sections. Therefore, in next few paragraph we will focus on how to find the particular solution for $Y(t)$ of the nonhomogeneous Eq. (C.61). There are two methods that can be considered as follows:

1. Undetermined Coefficients
2. Variation of Parameters.

There some advantages and some possible shortcoming for each of these above methods [6].

C.3.4.1 Method of Undetermined Coefficients

The method of undetermined coefficients requires that we make an initial assumption about the form of the particular solution $Y(t)$, but with the coefficients left unspecified. We then substitute the assumed expression into Eq. (C.60) and attempt to determine the coefficients so as to satisfy that equation. If we are successful, then we have found a solution of the differential Eq. (C.60) and can use it for the particular solution $Y(t)$. If we cannot determine the coefficients, then this means that there is no solution of the form that we assumed. In this case we may modify the initial assumption and try again.

The main advantage of the method of undetermined coefficients is that it is straightforward to execute once the assumption is made as to the form of $Y(t)$. Its major limitation is that it is useful primarily for equations for which we can easily write down the correct form of the particular solution in advance. For this reason, this method is usually used only for problems in which the homogeneous equation has constant coefficients and the nonhomogeneous term is restricted to a relatively small class of functions. In particular, we consider only nonhomogeneous terms that consist of polynomials, exponential functions, sines, and cosines. Despite this limitation, the method of undetermined coefficients is useful for solving many problems that have important applications. However, the algebraic details may become tedious and a computer algebra system can be very helpful in practical applications. We will illustrate the method of undetermined coefficients by several simple examples and then summarize some rules for using it [5].

Example 1: Find a particular solution of $y'' + 3y' - 4y = 3e^{2t}$

Solution: Looking for a function Y such that the combination $Y''(t) - 3Y'(t) - 4Y(t)$ is equal to $3e^{2t}$. Since the exponential function reproduce itself through differentiation, the most possible way to achieve the desired result is to assume $Y(t)$ is some multiple of e^{2t}, that is, $Y(t) = Ae^{2t}$ where the coefficient A is yet to be determined. To find A we calculate

$$Y''(t) = 2Ae^{2t} \quad Y''(t) = 4Ae^{2t}$$

and substituting for y, y', and y'' in the differential equation of example we have;

$$(4A - 6A - 4A)e^{2t} = 3e^{t}$$

This result that $A = -1/2$. Thus a particular solution is $A = -\frac{1}{2}e^{2t}$.

Example 2: Find a particular solution of $y'' - 3y' - 4y = 2 \sin t$.

Solution: Let us assume $Y(t) = A \sin t$, then we have the following steps:

$$-5A \sin t - 3A \cos t = 2 \sin t$$

$$(2 + 5A) \sin t + 3A \cos t = 0 \tag{C.64}$$

The functions $\sin t$ and $\cos t$ are linearly independent, so Eq. (C.64) can hold on an interval only if the coefficients $2 + 5A$ and $3A$ are both zero. These contradictory requirements mean that there is no choice of the constant A that makes Eq. (C.64) true for all t. Thus we conclude that our assumption concerning $Y(t)$ is inadequate. The appearance of the cosine term in Eq. (C.64) suggests that we modify our original assumption to include a cosine term in $Y(t)$, that is,

$$Y(t) = A\sin t + B\cos t$$

where A and B are to be determined. Then

$$Y'(t) = A\cos t - B\sin t \quad Y''(t) = -A\sin t - B\cos t$$

Substituting for y, y' and y'' we have

$$(-A + 3B - 4A)\sin t + (-B - 3A - 4B)\cos t = 2\sin t$$

$$A = -5/17 \quad B = 3/17$$

$$Y(t) = -\frac{5}{17}\sin t + \frac{3}{17}\cos t$$

The method illustrated in the preceding examples can also be used when the right side of the equation is a polynomial. Thus, to find a particular solution of $y'' - 3y' - 4y = 4t^2 - 1$ we initially assume that $Y(t)$ is a polynomial of the same degree as the nonhomogeneous term, that is, $Y(t) = At^2 + Bt + C$.

In summary we can have the following table for this type of method of finding particular solution of a nonhomogeneous differential equations.

Summary

1. If nonhomogeneous term $g(t)$ is an exponential $e^{\alpha t}$ function then $Y(t)$ is proportional to the same exponential function.

$$y'' + 3y' - 4y = 3e^{2t} \quad \text{Then} \quad Y(t) = Ae^{2t}$$

2. If nonhomogeneous term $g(t)$ is $\sin\beta t$ or $\cos\beta t$, then $Y(t)$ is a linear combination of $\sin\beta t$ and $\cos\beta t$.

$$y'' - 3y' - 4y = 2\sin t \quad \text{Then} \quad Y(t) = A\sin t + B\cos t$$

3. If nonhomogeneous term $g(t)$ is a polynomial, and then assume $Y(t)$ is a polynomial of same degree.

$$y'' - 3y' - 4y = 4t^2 - 1 \quad \text{Then} \quad Y(t) = At^2 + Bt + C$$

4. If nonhomogeneous term $g(t)$ is a product of any two, or all three, of these types of functions, as

$$y'' - 3y' - 4y = -8e^t\cos t2t \quad \text{Then} \quad Y(t) = Ae^t\cos 2t + Be^t\sin 2t$$

We encourage that reader of this book to refer to Boyce and DiPrima of reference [5] in this appendix

C.3.4.2 Method of Variation of Parameters

The procedure that we used in Sect. 1.8 to find a particular solution of a linear First-Order Differential Equation on an interval is applicable to linear higher-order equations as well. To adapt a method such as Variation of Parameters to a Linear Second-Order Differential Equation such as Eq. (C.60) is a question that we need to address. We want to know if this method can be applied effectively to an arbitrary equation such Eq. (C.60) which is reflecting below again

$$y'' + p(t)y' + q(t)y = g(t) \tag{C.65}$$

where p, q, and g are given continuous functions. As we have learned in previous section the general solution of homogeneous part of Eq. (C.65) which is present below is give by Eq. (C.67) in form Complementary or Characteristic form of Eq. (C.65)

$$y'' + p(t)y' + q(t)y = 0 \tag{C.66}$$

$$y_c(t) = c_1 y_1(t) + c_2 y_2(t) \tag{C.67}$$

Based on this major assumption it is shown how to solve Eq. (C.66) providing that it has constant coefficients. If the coefficients of Eq. (C.67) are function of variable t, then we have to use Series Solutions of Second-Order Linear Equations methods in order to find $y_c(t)$.

Now we have to find a particular solution of $y_p(t)$ that satisfies nonhomogeneous part of Eq. (C.65) as well. We seek a solution of the form

$$y_p(t) = u_1(t)y_1(t) + u_2(t)y_2(t) \tag{C.68}$$

Now we have to identify $u_1(t)$ and $u_2(t)$ that Eq. (C.68) becomes solution of nonhomogeneous part of Eq. (C.65) rather than homogeneous Eq. (C.66). Thus we differentiate Eq. (C.68) to obtain the following;

$$y_p'(t) = u_1'(t)y_1(t) + u_1(t)y_1'(t) + u_2'(t)y_2(t) + u_2(t)y_2'(t) \tag{C.69}$$

$$y_p''(t) = u_1'(t)y_1'(t) + u_1(t)y_1''(t) + u_2'(t)y_2'(t) + u_2(t)y_2''(t) \tag{C.70}$$

Considering that we set the terms involving $u_1'(t)$ and $u_2'(t)$ Eq. (C.69) equal to zero; that is we require that

$$u_1'(t)y_1(t) + u_2'(t)y_2(t) = 0 \tag{C.71}$$

Now we substitute for y, y', and y'' in Eq. (C.65) from Eqs. (C.68), (C.69), and (C.70), respectively. After rearranging the terms in the resulting equation we find that

$$\begin{aligned} u_1(t)\left[y_1''(t) + p(t)y_1'(t) + q(t)y_1(t)\right] \\ + u_2(t)\left[y_2''(t) + p(t)y_2'(t) + q(t)y_2(t)\right] \\ + u_1'(t)y_1'(t) + u_2'(t)y_2'(t) = g(t) \end{aligned} \tag{C.72}$$

Each of the expressions in square brackets in Eq. (C.72) is zero because both y_1 and y_2 are solutions of the homogeneous Eq. (C.66). Therefore Eq. (C.72) reduces to

$$u_1'(t)y_1'(t) + u_2'(t)y_2'(t) = g(t) \tag{C.73}$$

By Cramer's rule, the solution of the system

$$\begin{cases} u_1'(t)y_1(t) + u_2'(t)y_2(t) \\ \\ u_1'(t)y_1'(t) + u_2'(t)y_2'(t) = g(t) \end{cases} \quad \text{or} \quad \begin{cases} y_1(t)u_1'(t) + y_2(t)u_2'(t) = 0 \\ \\ y_1'(t)u_1'(t) + y_2'(t)u_2'(t) = g(t) \end{cases} \tag{C.74}$$

can be expressed in terms of determinants:

$$u_1'(t) = \frac{W_1}{W} = -\frac{y_2(t)g(t)}{W} \quad \text{and} \quad u_2'(t) = \frac{W_2}{W} = -\frac{y_1(t)g(t)}{W} \tag{C.75}$$

where $W(y_1, y_2)$ is the Wronskian of $y_1(t)$ and $y_2(t)$. Note that division by $W(y_1, y_2)$ is permissible since $y_1(t)$ and $y_2(t)$ are a fundamental set of solutions, and therefore Wronskian is nonzero.

$$W(y_1, y_2) = \begin{vmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{vmatrix} \quad W_1 = \begin{vmatrix} 0 & y_2(t) \\ g(t) & y_2'(t) \end{vmatrix} \quad W_2 = \begin{vmatrix} y_1(t) & 0 \\ y_1'(t) & g(t) \end{vmatrix} \tag{C.76}$$

By integrating Eq. (C.75) we find the desired functions $u_1(t)$ and as follows:

$$u_1(t) = -\int \frac{y_2(t)g(t)}{W(y_1, y_2)}dt + c_1 \quad u_2(t) = \int \frac{y_1(t)g(t)}{W(y_1, y_2)}dt + c_2 \tag{C.77}$$

Finally, substituting the results of Eq. (C.77) into Eq. (C.68) we obtain the general solution of Eq. (C.65). We can summarize the result as a following theorem [5].

Theorem III
If the functions p, q, and g are continuous on an interval I, and if the functions y_1 and y_2 are linearly independent solutions of the homogeneous Eq. (C.66) corresponding to the nonhomogeneous Eq. (C.65),

$$y'' + p(t)y' + q(t)y = g(t)$$

Then a particular solution of Eq. (C.65) is given as

$$Y(t) = -y_1(t)\int \frac{y_2(t)g(t)}{W(y_1, y_2)}dt + y_2(t)\int \frac{y_1(t)g(t)}{W(y_1, y_2)}dt \tag{C.78}$$

and the general solution is

$$y = c_1 y_1(t) + c_2 y_2(t) + Y(t) \tag{C.79}$$

as prescribed by *Theorem II*.

Example 1: Solve $y'' - 4y' + 4y = (x+1)e^{2x}$

Solution: The auxiliary equation of complementary term is given by $m^2 - 4m + 4 = (m-2)^2 = 0$, which has repeating roots therefore we have $y = c_1 e^{2x} + c_1 x e^{2x}$. With the identification $y_1 = e^{2x}$ and $y_2 = xe^{2x}$, we next compute the Wronskian [4]:

$$W\left(e^{2x}, xe^{2x}\right) = \begin{vmatrix} e^{2x} & xe^{2x} \\ 2xe^{2x} & 2xe^{2x} + e^{2x} \end{vmatrix} = e^{4x}$$

Since the given differential equation is already in the form of Eq. (C.65) (that is, the coefficient of y'' is 1), we identify $f(x) = (x+1)e^{2x}$. From Eq. (C.76) we obtain

$$W_1 = \begin{vmatrix} 0 & xe^{2x} \\ (x+1)e^{2x} & 2xe^{2x} + e^{2x} \end{vmatrix} = -(x+1)xe^{4x}$$

$$W_2 = \begin{vmatrix} e^{2x} & 0 \\ 2e^{2x} & (x+1)e^{2x} \end{vmatrix} = (x+1)xe^{4x}$$

And so from Eq. (C.75) we have

$$u_1' = -\frac{(x+1)xe^{4x}}{e^{4x}} = -x^2 - x \quad u_2' = -\frac{(x+1)xe^{4x}}{e^{4x}} = x+1$$

It follows that $u_1 = -\frac{1}{3}x^3 - \frac{1}{2}x^2$ and $u_2 = \frac{1}{2}x^2 + x$. Hence

$$y_p = \left(-\frac{1}{3}x^3 - \frac{1}{2}x^2\right)e^{2x} + \left(\frac{1}{2}x^2 + x\right)xe^{2x} = \frac{1}{6}x^3e^{2x} + \frac{1}{2}x^2e^{2x}$$

$$\text{and} \quad y = y_c + y_p = c_1e^{2x} + c_2xe^{2x} + \frac{1}{6}x^3e^{2x} + \frac{1}{2}x^2e^{2x}$$

Example 2: Solve $4y'' + 36y = \csc\ 3x$

Solution: Here we are not spending time to explaining the details of solution and we leave that to reader but show all the steps

$$y'' + 9y = \frac{1}{4}\csc\ 3x$$

$$m^2 + 9 = 0 \quad \begin{cases} m_1 = 3i \\ m_2 = -3i \end{cases} \quad \text{Complex roots for complementary function}$$

$$y_c = c_1 \cos\ 3x + c_2\ \sin\ 3x$$

$$\begin{cases} y_1 = \cos 3x \\ y_2 = \sin 3x \end{cases} \quad \text{and} \quad f(x) = \frac{1}{4}\csc 3x$$

The Wronskian [5]:

$$W(\cos 3x, \sin 3x) = \begin{vmatrix} \cos 3x & \sin 3x \\ -3\sin 3x & 3\cos 3x \end{vmatrix} = 3$$

$$W_1 = \begin{vmatrix} 0 & \sin 3x \\ \frac{1}{4}\csc 3x & 3\cos 3x \end{vmatrix} = \frac{1}{4} \quad W_2 = \begin{vmatrix} \cos 3x & 0 \\ -3\sin 3x & \frac{1}{4}\csc 3x \end{vmatrix} = \frac{1}{4}\frac{\cos 3x}{\sin 3x}$$

Integrating, we have

$$u_1' = \frac{W_1}{W} = -\frac{1}{4} \quad u_2' = \frac{W_2}{W} = \frac{1}{4}\frac{\cos 3x}{\sin 3x}$$

$$\text{Then} \quad u_1 = -\frac{1}{12}x \quad \text{and} \quad u_2 = \frac{1}{36}\ln|\sin 3x|$$

Thus the particular solution is $y_p = -\frac{1}{2}x\cos 3x + \frac{1}{36}(\sin 3x)\ln|\sin 3x|$
The general solution say between interval $(0, \pi/6)$ is [5].

$$y = y_c + y_p = c_1\cos 3x + c_2\sin 3x - \frac{1}{12}x\cos 3x + \frac{1}{36}(\sin 3x)\ln|\sin 3x|$$

Example 3: Solve $y'' - y = \frac{1}{x}$

Solution: The auxiliary equation $m^2 - 1 = 0$ yield $m_1 = -1$ and $m_2 = 1$ [4]. Therefore $y_c = c_1e^x + c_2e^{-x}$. Now $W(e^x, e^{-x}) = -2$ and

$$u_1' = -\frac{e^{-x}(1/x)}{-2} \quad u_1 = \frac{1}{2}\int_{x_0}^{x}\frac{e^{-t}}{t}dt$$

$$u_2' = -\frac{e^{-x}(1/x)}{-2} \quad u_2 = -\frac{1}{2}\int_{x_0}^{x}\frac{e^{-t}}{t}dt$$

Due to fact that forgoing integrals are nonelementary type, we are forced to write

$$y_p = \frac{1}{2}e^x\int_{x_0}^{x}\frac{e^{-t}}{t}dt - \frac{1}{2}e^{-x}\int_{x_0}^{x}\frac{e^{-t}}{t}dt$$

and

$$y = y_c + y_p = c_1e^x + c_2e^{-x} + \frac{1}{2}e^x\int_{x_0}^{x}\frac{e^{-t}}{t}dt - \frac{1}{2}e^{-x}\int_{x_0}^{x}\frac{e^{-t}}{t}dt$$

C.4 Eigenvalues and Eigenvectors

We need a quick review of linear algebra we need to get this section. Without this we will not be able to do any of the differential equations work that is in this section and appendix.

So let's start with the following. If we multiply an $n \times n$ matrix by an $n \times 1$ vector we will get a new $n \times 1$ vector back. In other words,

$$\mathrm{A}\vec{\eta} = \vec{\mathrm{y}}$$

What we want to know is if it is possible for the following to happen. Instead of just getting a brand new vector out of the multiplication is it is possible instead to get the following:

$$A\vec{\eta} = \lambda\vec{\eta} \tag{C.80}$$

In other words is it possible, at least for certain λ and $\vec{\eta}$, to have matrix multiplication be the same as just multiplying the vector by a constant?. Of course, we probably wouldn't be talking about this if the answer was no. So, it is possible for

this to happen, however, it won't happen for just any value of λ or $\vec{\eta}$. If we do happen to have a λ and $\vec{\eta}$ for which this works (and they will always come in pairs) then we call λ an **eigenvalue** of A and $\vec{\eta}$ an **eigenvector** of A.

So, how do we go about find the eigenvalues and eigenvectors for a matrix? Well first notice that that if $\vec{\eta} = \vec{0}$ then Eq. (C.14) is going to be true for any value of λ and so we are going to make the assumption that. With that out of the way let's rewrite Eq. (C.14) a little.

$$\begin{aligned} A\vec{\eta} - \lambda\vec{\eta} &= \vec{0} \\ A\vec{\eta} - \lambda I_n\vec{\eta} &= \vec{0} \\ (A - \lambda I_n)\vec{\eta} &= \vec{0} \end{aligned}$$

Notice that before we factored out the $\vec{\eta}$ we added in the appropriately sized identity matrix. This is equivalent to multiplying things by a one and so doesn't change the value of anything. We needed to do this because without it we would have had the difference of a matrix, A, and a constant, λ, and this can't be done. We now have the difference of two matrices of the same size which can be done.

So, with this rewrite we see that

$$(A - \lambda I_n)\vec{\eta} = \vec{0} \tag{C.81}$$

is equivalent to Eq. (C.80). In order to find the eigenvectors for a matrix we will need to solve a homogeneous system. Recall the fact from the previous section that we know that we will either have exactly one solution $\left(\vec{\eta} = \vec{0}\right)$ or we will have infinitely many nonzero solutions. Since we've already said that don't want $\vec{\eta} = \vec{0}$ this means that we want the second case.

Knowing this will allow us to find the eigenvalues for a matrix. Recall from this fact that we will get the second case only if the matrix in the system is singular. Therefore we will need to determine the values of λ for which we get,

$$\det(A - \lambda I) = 0$$

Once we have the eigenvalues we can then go back and determine the eigenvectors for each eigenvalue. Let's take a look at a couple of quick facts about eigenvalues and eigenvectors.

Fact
If A is an $n \times n$ matrix then $\det(A - \lambda I) = 0$ is an n^{th} degree polynomial. This polynomial is called the **Characteristic Polynomial**.

To find eigenvalues of a matrix all we need to do is solve a polynomial. That's generally not too bad provided we keep n small. Likewise this fact also tells us that for an $n \times n$ matrix, A, we will have n eigenvalues if we include all repeated eigenvalues.

Fact
If $\lambda_1, \lambda_2 \ldots .., \lambda_n$ is the complete list of eigenvalues for A (including all repeated eigenvalues) then,

1. If λ occurs only once in the list then we call λ **simple.**
2. If λ occurs $k > I$ times in the list then we say that λ has **multiplicity** $\boldsymbol{k}$.
3. If $\lambda_1, \lambda_2 \ldots ., \lambda_n \ (k \leq n)$ are the simple eigenvalues in the list with corresponding eigenvectors $\vec{\eta}^{(1)}, \vec{\eta}^{(2)} \ldots ., \vec{\eta}^{(k)}$ then the eigenvectors are all linearly independent.
4. If λ is an eigenvalue of $k > I$ then λ will have anywhere from 1 to k linearly independent eigenvectors.

The usefulness of these facts will become apparent when we get back into differential equations since in that work we will want linearly independent solutions.

For those readers that have access to website, I encourage you go to the following link to go to and get more information on the Paul's Online Math Notes http://tutorial.math.lamar.edu/

Let's work a couple of examples now to see how we actually go about finding eigenvalues and eigenvectors.

Example 1: Find the eigenvalues and eigenvectors of the following matrix

$$A = \begin{pmatrix} 2 & 7 \\ -1 & -6 \end{pmatrix}$$

Solution:
The first thing that we need to do is find the eigenvalues. That means we need the following matrix,

$$A - \lambda I = \begin{pmatrix} 2 & 7 \\ -1 & -6 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2-\lambda & 7 \\ -1 & -6-\lambda \end{pmatrix}$$

In particular we need to determine where the determinant of this matrix is zero.

$$\det(A - \lambda I) = (2-\lambda)(-6-\lambda) + 7 = \lambda^2 + 4\lambda - 5 = (\lambda + 5)(\lambda - 1)$$

So, it looks like we will have two simple eigenvalues for this matrix, $\lambda_1 = -5$ and $\lambda_2 = 1$. We will now need to find the eigenvectors for each of these. Also

(continued)

note that according to the fact above, the two eigenvectors should be linearly independent.

To find the eigenvectors we simply plug in each eigenvalue into Eq. (C.15) and solve. So, let's do that.

$$\lambda_1 = -5 :$$

In this case we need to solve the following system

$$\begin{pmatrix} 7 & 7 \\ -1 & -1 \end{pmatrix} \vec{\eta} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Recall that officially to solve this system we use the following augmented matrix.

$$\begin{pmatrix} 7 & 7 & 0 \\ -1 & -1 & 0 \end{pmatrix} \overset{\frac{1}{7}R_1+R_2}{\Rightarrow} \begin{pmatrix} 7 & 7 & 0 \\ -1 & -1 & 0 \end{pmatrix}$$

Upon reducing down we see that we get a single equation

$$7\eta_1 + 7\eta_2 = 0 \quad \Rightarrow \quad \eta_1 = -\eta_2$$

that will yield an infinite number of solutions. This is expected behavior. Recall that we picked the eigenvalues so that the matrix would be singular and so we would get infinitely many solutions.

Notice as well that we could have identified this from the original system. This won't always be the case, but in the 2 x 2 case we can see from the system that one row will be a multiple of the other and so we will get infinite solutions. From this point on we won't be actually solving systems in these cases. We will just go straight to the equation and we can use either of the two rows for this equation.

Now, let's get back to the eigenvector, since that is what we were after. In general then the eigenvector will be any vector that satisfies the following,

$$\vec{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} -\eta_1 \\ \eta_2 \end{pmatrix} \quad \eta_2 \neq 0$$

We really don't want a general eigenvector however so we will pick a value for η_2 to get a specific eigenvector. We can choose anything (except $\eta_2 = 0$), so pick something that will make the eigenvector "nice". Note as well that

(continued)

since we've already assumed that the eigenvector is not zero we must choose a value that will not give us zero, which is why we want to avoid $\eta_2 = 0$ in this case. Here's the eigenvector for this eigenvalue.

$$\vec{\eta}^{(1)} = \begin{pmatrix} -1 \\ \\ 1 \end{pmatrix} \quad \text{using} \quad \eta_2 = 1$$

Now we get to do this all over again for the second eigenvalue.

$$\lambda_2 = 1 :$$

We'll do much less work with this part than we did with the previous part. We will need to solve the following system.

$$\begin{pmatrix} 1 & 7 \\ -1 & -7 \end{pmatrix} \vec{\eta} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Clearly both rows are multiples of each other and so we will get infinitely many solutions. We can choose to work with either row. We'll run with the first because to avoid having too many minus signs floating around. Doing this gives us,

$$\eta_1 + 7\eta_2 = 0 \quad \eta_1 = -7\eta_2$$

Note that we can solve this for either of the two variables. However, with an eye towards working with these later on let's try to avoid as many fractions as possible. The eigenvector is then,

$$\vec{\eta}^{(2)} = \begin{pmatrix} -7 \\ \\ 1 \end{pmatrix} \quad \text{using} \quad \eta_2 = 1$$

$$\vec{\eta}^{(2)} = \begin{pmatrix} -7 \\ \\ 1 \end{pmatrix} \quad \text{using} \quad \eta_2 = 1$$

Summarizing we have,

(continued)

$$\lambda_1 = -5 \quad \vec{\eta}^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\lambda_2 = 1 \quad \vec{\eta}^{(2)} = \begin{pmatrix} -7 \\ 1 \end{pmatrix}$$

Note that the two eigenvectors are linearly independent as predicted.

C.5 Eigenvalues and Eigenfunctions

For a given square matrix, A, if we could find values of λ for which we could find nonzero solutions, i.e. $\vec{x} \neq \vec{0}$, to,

$$A\vec{x} = \lambda\vec{x}$$

then we called λ an eigenvalue of A and was its corresponding eigenvector. It's important to recall here that in order for λ to be an eigenvalue then we had to be able to find nonzero solutions to the equation. In order to show the relationship between this topic and Boundary Value problem (BVP) we show the two following examples and expand on them. In these two examples we solve homogeneous (and that is very important to notice) BVP's in the form of,

$$y'' + \lambda y = 0 \quad y(0) = 0 \quad y(2\pi) = 0 \tag{C.82}$$

Example 1: Solve the following BVP

$$y'' + 4y = 0 \quad y(0) = 0 \quad y(2\pi) = 0$$

Solution: Here the general solution is,

$$y(x) = c_1 \cos(2x) + c_2 \sin(2x)$$

Applying the boundary conditions gives

$$0 = y(0) = c_1$$

$$0 = y(2\pi) = c_1$$

So c_2 is arbitrary and the solution is $y(x) = c_2 \sin(2x)$ and in this case we will get infinitely many solutions.

Example 2: Solve the following BVP

$$y'' + 3y = 0 \quad y(0) = 0 \quad y(2\pi) = 0$$

Solution: The general solution in this case is

$$y(x) = c_1 \cos\left(\sqrt{3}x\right) + c_2 \sin\left(\sqrt{3}x\right)$$

Applying the boundary conditions gives

$$0 = y(0) = c_1$$

$$0 = y(2\pi) = c_2 \sin\left(\sqrt{3}x\right) \quad \Rightarrow \quad c_2 = 0$$

In this case we found both constants to be zero and so the solution is, $y(x) = 0$

In Example 1 we had $\lambda = 4$ and we found nontrivial (i.e. nonzero) solutions to the BVP. In Example 2 we used $\lambda = 3$ and the only solution was the trivial solution (i.e. $y(x) = 0$). So, this homogeneous BVP (recall this also means the boundary conditions are zero) seems to exhibit similar behavior to the behavior in the matrix equation above. There are values of λ that will give nontrivial solutions to this BVP and values of λ that will only admit the trivial solution.

So, for those values of λ that give nontrivial solutions we'll call λ an **eigenvalue** for the BVP and the nontrivial solutions will be called Eigenfunctions for the BVP corresponding to the given eigenvalue. We now know that for the homogeneous BVP given in Eq. (C.82), $\lambda = 4$ is an eigenvalue (with eigenfunctions $y(x) = c_2 \sin(2x)$) and that $\lambda = 3$ is not an eigenvalue.

Eventually we'll try to determine if there are any other eigenvalues for Eq. (C.82), however before we do that let's comment briefly on why it is so important for the BVP to homogeneous in this discussion. In Example 3 and Example 4 down below we solve the homogeneous differential equations.

$$y'' + \lambda y = 0$$

with two different nonhomogeneous boundary conditions in the form,

$$y(0) = a \quad y(2\pi) = b$$

Example 3: Solve the following BVP

$$y'' + 4y = 0 \quad y(0) = -2 \quad y(2\pi) = -2$$

Solution: We have a general solution of the form as follows:

$$y(x) = c_1 \cos(2x) + c_2 \sin(2x)$$

Upon applying the boundary conditions we get

$$-2 = y(0) = c_1$$

$$-2 = y(2\pi) = c_1$$

So in this case, unlike previous example, both boundary conditions tell us that we have to have $c_1 = -2$ and neither one of them tell us anything about c_2. Remember however that all we're asking for is a solution to the differential equation that satisfies the two given boundary conditions and the following function will do that

$$y(x) = -2 \cos(2x) + c_2 \sin(2x)$$

In other words, regardless of the value of c_2 we get a solution and so, in this case we get infinitely many solutions to the boundary value problem.

Example 4: Solve the following BVP.

$$y'' + 4y = 0 \quad y(0) = -2 \quad y(2\pi) = 3$$

Solution: Again we have the following general solution as follows:

$$y(x) = c_1 \cos(2x) + c_2 \sin(2x)$$

This time the boundary conditions give us,

$$-2 = y(0) = c_1$$

$$3 = y(2\pi) = c_1$$

In this case we have a set of boundary conditions each of which requires a different value of c_1 in order to be satisfied. This, however, is not possible and so in this case have no solution.

In these two examples we saw that by simply changing the value of a and/or b we were able to get either nontrivial solutions or to force no solution at all. In the discussion of eigenvalues/eigenfunctions we need solutions to exist and the only way to assure this behavior is to require that the boundary conditions also be homogeneous. In other words, we need for the BVP to be homogeneous. There is one final topic that we need to discuss before we move into the topic of eigenvalues and eigenfunctions and this is more of a notational issue that will help us with some of the work that we'll need to do.

Let's suppose that we have a second order differential equations and its characteristic polynomial has two real, distinct roots and that they are in the form

$$r_1 = \alpha \quad r_2 = -\alpha$$

Then we know that the solution is,

$$y(x) = c_1 e^{r_1 x} + c_2 e^{r_1 x} = c_1 e^{\alpha x} + c_2 e^{-\alpha x}$$

While there is nothing wrong with this solution let's do a little rewriting of this. We'll start by splitting up the terms as follows,

$$\begin{aligned} y(x) &= c_1 e^{\alpha x} + c_2 e^{-\alpha x} \\ &= \frac{c_1}{2} e^{\alpha x} + \frac{c_1}{2} e^{\alpha x} + \frac{c_2}{2} e^{-\alpha x} + \frac{c_2}{2} e^{-\alpha x} \end{aligned}$$

Now we'll add/subtract the following terms (note we're "mixing" the c_i and $\pm\alpha$ up in the new terms) to get,

$$y(x) = \frac{c_1}{2} e^{\alpha x} + \frac{c_1}{2} e^{\alpha x} + \frac{c_2}{2} e^{-\alpha x} + \frac{c_2}{2} e^{-\alpha x} + \left(\frac{c_1}{2} e^{-\alpha x} - \frac{c_1}{2} e^{-\alpha x}\right) + \left(\frac{c_2}{2} e^{\alpha x} - \frac{c_2}{2} e^{\alpha x}\right)$$

Next, rearrange terms around a little,

$$y(x) = \frac{1}{2}\left(c_1 e^{\alpha x} + c_1 e^{-\alpha x} + c_2 e^{\alpha x} + c_2 e^{-\alpha x}\right) + \frac{1}{2}\left(c_1 e^{\alpha x} - c_1 e^{-\alpha x} - c_2 e^{\alpha x} + c_2 e^{-\alpha x}\right)$$

Finally, the quantities in parenthesis factor and we'll move the location of the fraction as well. Doing this, as well as renaming the new constants we get,

$$\begin{aligned} y(x) &= (c_1 + c_2)\frac{e^{\alpha x} + e^{-\alpha x}}{2} + (c_1 - c_2)\frac{e^{\alpha x} - e^{-\alpha x}}{2} \\ &= C_1 \frac{e^{\alpha x} + e^{-\alpha x}}{2} + C_2 \frac{e^{\alpha x} - e^{-\alpha x}}{2} \end{aligned}$$

All this work probably seems very mysterious and unnecessary. However there really was a reason for it. In fact you may have already seen the reason, at least in part. The two "new" functions that we have in our solution are in fact two of the hyperbolic functions. In particular,

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \quad \sinh(x) = \frac{e^x - e^{-x}}{2}$$

So, another way to write the solution to a second order differential equation whose characteristic polynomial has two real, distinct roots in the form $r_1 = \alpha, r_2 = -\alpha$ is,

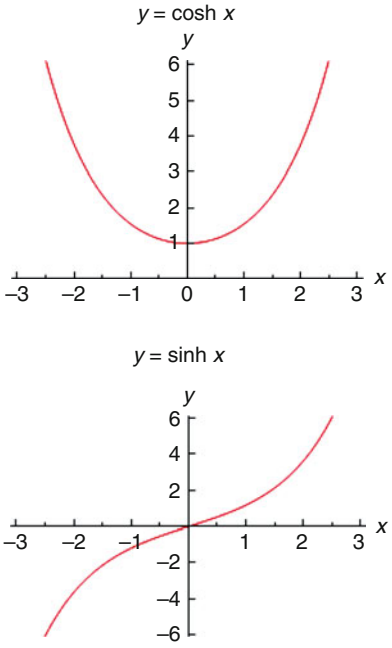
$$y(x) = C_1\cosh(\alpha x) + C_2\sinh(\alpha x)$$

Having the solution in this form for some (actually most) of the problems we'll be looking will make our life a lot easier. The hyperbolic functions have some very nice properties that we can (and will) take advantage of.

First, since we'll be needing them later on, the derivatives are,

$$\frac{d}{dx}(\cosh(x)) = \sinh(x) \quad \frac{d}{dx}(\sinh(x)) = \cosh(x)$$

Next let's take a quick look at the graphs of these functions.



Note that $\cosh(0) = 1$ and $\sinh(0) = 0$. Because we'll often be working with boundary conditions at $x = 0$ these will be useful evaluations.

Next, and possibly more importantly, let's notice that $\cosh(x) > 0$ for all x and so the hyperbolic cosine will never be zero. Likewise, we can see that $\sinh(x) = 0$ only if $x = 0$. We will be using both of these facts in some of our work so we shouldn't forget them.

Now let's work an example to see how we go about finding eigenvalues/eigenfunctions for a BVP.

Example 5
Find all the eigenvalues and eigenfunctions for the following BVP.

$$y'' + \lambda y = 0 \quad y(0) = 0 \quad y(2\pi) = 0$$

Solution
We started off this section looking at this BVP and we already know one eigenvalue $(\lambda = 4)$ and we know one value of λ that is not an eigenvalue $(\lambda = 3)$. As we go through the work here we need to remember that we will get an eigenvalue for a particular value of λ if we get non-trivial solutions of the BVP for that particular value of λ.

In order to know that we've found all the eigenvalues we can't just start randomly trying values of λ to see if we get non-trivial solutions or not. Luckily there is a way to do this that's not too bad and will give us all the eigenvalues/eigenfunctions. We are going to have to do some cases however. The three cases that we will need to look at are: $\lambda > 0, \lambda = 0, \lambda < 0$ and . Each of these cases gives a specific form of the solution to the BVP which we can then apply the boundary conditions to see if we'll get non-trivial solutions or not. So, let's get started on the cases.

$$\lambda > 0$$

In this case the characteristic polynomial we get from the differential equation is,

$$r^2 + \lambda = 0 \quad \Rightarrow \quad r_{1,2} = \pm\sqrt{-\lambda}$$

In this case since we know that $\lambda > 0$ these roots are complex and we can write them instead as, $r_{1,2} = \pm i\sqrt{\lambda}$ and the general solution to the differential equation is then,

$$y(x) = c_1 \cos\left(\sqrt{\lambda}x\right) + c_2 \sin\left(\sqrt{\lambda}x\right)$$

Applying the first boundary condition gives us,

$$0 = y(0) = c_1$$

So, taking this into account and applying the second boundary condition we get,

(continued)

$$0 = y(2\pi) = c_2 \sin\left(2\pi\sqrt{\lambda}\right)$$

This means that we have to have one of the following,

$$c_2 = 0 \quad \text{or} \quad \sin\left(2\pi\sqrt{\lambda}\right) = 0$$

However, recall that we want non-trivial solutions and if we have the first possibility we will get the trivial solution for all values of $\lambda > 0$. Therefore let's assume that $c_2 \neq 2$. This means that we have,

$$\sin\left(2\pi\sqrt{\lambda}\right) = 0 \quad \Rightarrow \quad 2\pi\sqrt{\lambda} = n\pi \quad n = 1, 2, 3, \cdots$$

In other words, taking advantage of the fact that we know where sine is zero we can arrive at the second equation. Also note that because we are assuming that $\lambda > 0$ we know that $2\pi\sqrt{\lambda} > 0$ and so n can only be a positive integer for this case. Now all we have to do is solve this for λ and we'll have all the positive eigenvalues for this BVP. The positive eigenvalues are then,

$$\lambda_n = \left(\frac{n}{2}\right)^2 = \frac{n^2}{4} \quad n = 1, 2, 3, \cdots$$

and the eigenfunctions that correspond to these eigenvalues are

$$y_n(x) = \sin\left(\frac{n\pi}{2}\right) \quad n = 1, 2, 3, \cdots$$

Note that we subscripted an n on the eigenvalues and eigenfunctions to denote the fact that there is one for each of the given values of n. Also note that we dropped the c_2 on the eigenfunctions. For eigenfunctions we are only interested in the function itself and not the constant in front of it and so we generally drop that.

Let's now move into the second case.

$$\lambda = 0$$

In this case the BVP becomes,

$$y'' = 0 \quad y(0) = 0 \quad y(2\pi) = 0$$

and integrating the differential equation a couple of times gives us the general solution,

(continued)

$$y(x) = c_1 + c_2 x$$

Applying the first boundary condition gives,

$$0 = y(0) = c_1$$

Applying the second boundary condition as well as the results of the first boundary condition gives,

$$0 = y(2\pi) = 2c_2\pi$$

Here, unlike the first case, we don't have a choice on how to make this zero. This will only be zero if $c_2 = 0$. Therefore, for this BVP (and that's important), if we have $\lambda = 0$ the only solution is the trivial solution and so $\lambda = 0$ cannot be an eigenvalue for this BVP.

Now let's look at the final case.

$$\lambda < 0$$

In this case the characteristic equation and its roots are the same as in the first case. So, we know that,

$$r_{1,2} = \pm\sqrt{-\lambda}$$

However, because we are assuming $\lambda < 0$ here these are now two real distinct roots and so using our work above for these kinds of real, distinct roots we know that the general solution will be,

$$y(x) = c_1\cosh\left(\sqrt{\lambda}x\right) + c_2\sinh\left(\sqrt{\lambda}x\right)$$

Note that we could have used the exponential form of the solution here, but our work will be significantly easier if we use the hyperbolic form of the solution here. Now, applying the first boundary condition gives,

$$0 = y(0) = c_1\cosh(0) + c_2\sinh(0) = c_1(1) + c_2(0) = c_1 \quad \Rightarrow \quad c_1 = 0$$

Applying the second boundary condition gives,

$$0 = y(2\pi) = c_2\sinh\left(2\pi\sqrt{-\lambda}\right)$$

(continued)

Because we are assuming $\lambda < 0$ we know that $2\pi\sqrt{-\lambda} \neq 0$ and so we also know that $\sinh\left(2\pi\sqrt{-\lambda}\right) \neq 0$. Therefore, much like the second case, we must have $c_2 = 0$.

So, for this BVP (again that's important), if we have $\lambda < 0$ we only get the trivial solution and so there are no negative eigenvalues.

In summary then we will have the following eigenvalues/eigenfunctions for this BVP.

$$\lambda_n = \frac{n^2}{4} y_n(x) = \sin\left(\frac{n\pi}{2}\right) n = 1, 2, 3, \cdots$$

Let us take a look at another example with slightly different boundary conditions.

Example 6
Find all the eigenvalues and eigenfunctions for the following BVP

$$y'' + \lambda y = 0 \quad y'(0) = 0 \quad y'(2\pi) = 0$$

Solution:
Here we are going to work with derivative boundary conditions. The work is pretty much identical to the previous example however so we won't put in quite as much detail here. We'll need to go through all three cases just as the previous example so let's get started on that.

$$\lambda > 0$$

The general solution to the differential equation is identical to the previous example and so we have

$$y(x) = c_1 \cos\left(\sqrt{\lambda}x\right) + c_2 \sin\left(\sqrt{\lambda}x\right)$$

Applying the first boundary condition gives us,

$$0 = y'(0) = \sqrt{\lambda}c_2 \quad \Rightarrow \quad c_2 = 0$$

Recall that we are assuming that $\lambda > 0$ here and so this will only be zero if $c_2 = 0$. Now, the second boundary condition gives us,

$$0 = y'(2\pi) = -\sqrt{\lambda}c_1 \sin\left(2\pi\sqrt{\lambda}\right)$$

(continued)

Recall that we don't want trivial solutions and that $\lambda > 0$ so we will only get non-trivial solution if we require that,

$$\sin\left(2\pi\sqrt{\lambda}\right) \quad \Rightarrow \quad 2\pi\sqrt{\lambda} \quad n = 1, 2, 3, \cdots$$

Solving for λ and we see that we get exactly the same positive eigenvalues for this BVP that we got in the previous example.

$$\lambda_n = \left(\frac{n}{2}\right)^2 = \frac{n^2}{4} n = 1, 2, 3, \cdots$$

The eigenfunctions that correspond to these eigenvalues however are,

$$y_n(x) = \cos\left(\frac{nx}{2}\right) \quad n = 1, 2, 3, \cdots$$

So, for this BVP we get cosines for eigenfunctions corresponding to positive eigenvalues.

Now the second case.

$$\lambda = 0$$

The general solution is, $y(x) = c_1 + c_2 x$. Applying the first boundary condition gives, $0 = y'(0) = c_2$. Using this result, the general solution is then, $y(x) = c_1$ and note that this will trivially satisfy the second boundary condition, $0 = y'(2\pi) = 0$. Therefore, unlike the first example, $\lambda = 0$ is an eigenvalue for this BVP and the eigenfunctions corresponding to this eigenvalue is, $y(x) = 1$. Again, note that we dropped the arbitrary constant for the eigenfunctions.

Finally let's take care of the third case.

$$\lambda < 0$$

The general solution here is,

$$y(x) = c_1 \cosh\left(\sqrt{-\lambda}x\right) + c_2 \sinh\left(\sqrt{-\lambda}x\right)$$

Applying the first boundary condition gives,

$$0 = y'(0) = \sqrt{-\lambda}c_1 \sinh(0) + \sqrt{-\lambda}c_2 \cosh(0) = \sqrt{-\lambda}c_2 \quad \Rightarrow \quad c_2 = 0$$

Applying the second boundary condition gives,

(continued)

$$0 = y'(2\pi)\sqrt{-\lambda}c_1\sinh\left(2\pi\sqrt{-\lambda}\right)$$

As with the previous example we again know that $2\pi\sqrt{-\lambda} \neq 0$ and so $\sinh\left(2\pi\sqrt{-\lambda}\right) \neq 0$. Therefore we must have $c_1 = 0$. So, for this BVP we again have no negative eigenvalues.

In summary then we will have the following eigenvalues/eigenfunctions for this BVP.

$$\lambda_n = \frac{n^2}{4} \qquad y_n(x) = \cos\left(\frac{nx}{2}\right) \quad n = 1, 2, 3, \cdots \quad \lambda_0 = 0 \qquad y_0(x) = 1$$

Notice as well that we can actually combine these if we allow the list of n's for the first one to start at zero instead of one. This will often not happen, but when it does we'll take advantage of it. So the "official" list of eigenvalues/eigenfunctions for this BVP is,

$$\lambda_n = \frac{n^2}{4} \qquad y_n(x) = \cos\left(\frac{nx}{2}\right) \qquad n = 1, 2, 3, \cdots$$

For those readers that have access to web site, I encourage you to go to the following link to go to and get more information on the Paul's Online Math Notes http://tutorial.math.lamar.edu/

C.6 Higher-Order Differential Equations

We now turn our attention to the solution of differential equations of order two or higher but we also remind you for the purpose of this book we do not need to know beyond second order differential equations (ODEs) behavior and their solution. The only equation we deal with is heat conduction and equation which is a second-order type with various Intial and Boundary conditions. Therefore readers that are interested to understand and learn more about ODEs beyond second order should refer to references [4] and [5] of this appendix.

C.7 Partial Differential Equation

This section discusses elementary partial differential equations in engineering and physical sciences. It is suited for courses whose titles include Fourier Series orthogonal functions, or boundary value problem. It may also be used in courses

on Green's Function or Transform Methods such as Laplace Transform, etc. (See Appendix E). Simple models (Heat Flow, Vibrating Strings and Membranes) are emphasized. Equations are formulated carefully from physical principles, motivating most mathematical topics. Standard topics such as the method of separation of variables, Fourier series, Laplace Transformation, and orthogonal functions are developed with considerable detail. In addition, there is variety of clearly presented topics, such as differentiation and integration of Fourier Series, Reverse Laplace Transformation, zeros of Sturm-Liouville Egienfunctions, Rayleigh quotient, Multidimensional Eigenvalue problems, Bessel Functions for vibration circular membrane, egienfunctions expansion for nonhomogeneous problems, Green's functions, Fourier and Laplace Transform solutions, method of characteristics, and numerical methods. Some optional advanced material of interest is also included (for example, asymptotic expansion of large eigenvalues, calculation of perturbed frequencies using the Fredholm alternative, and the dynamics of shock waves).

C.7.1 Definition

A partial differential equation (PDE) describes a relation between an unknown function and its partial derivatives. PDEs appear frequently in all areas of physics and engineering. Moreover, in recent years we have seen a dramatic increase in the use of PDEs in areas such as biology, chemistry, computer sciences (particularly in relation to image processing and graphics) and in economics (finance). In fact, in each area where there is an interaction between a numbers of independent variables, we attempt to define functions in these variables and to model a variety of processes by constructing equations for these functions. When the value of the unknown function(s) at a certain point depends only on what happens in the vicinity of this point, we shall, in general, obtain a PDE.

The general form of a PDE for a function $u(x_1, x_2, \ldots, x_n)$ is

$$F(x_1, x_2, \ldots, x_n, u, u_{x_1}, u_{x_2}, \ldots u_{x_{11}}, \ldots) = 0 \tag{C.83}$$

where $x_1, x_2, \ldots, x_n$ are the independent variables, u is the unknown function, and u_{x_i} denotes the partial derivative $\dfrac{\partial u}{\partial u_{x_i}}$. The equation is, in general, supplemented by additional conditions such as initial conditions (as we have often seen in the theory of ordinary differential equations (ODEs)) or boundary conditions.

The analysis of PDEs has many facets. The classical approach that dominated the nineteenth century was to develop methods for finding explicit solutions. Because of the immense importance of PDEs in the different branches of physics, every mathematical development that enabled a solution of a new class of PDEs was accompanied by significant progress in physics. Thus, the method of

characteristics invented by Hamilton led to major advances in optics and in analytical mechanics. The Fourier method enabled the solution of heat transfer and wave propagation, and Green's method was instrumental in the development of the theory of electromagnetism. The most dramatic progress in PDEs has been achieved in the last 50 years with the introduction of numerical methods that allow the use of computers to solve PDEs of virtually every kind, in general geometries and under arbitrary external conditions (at least in theory; in practice there are still a large number of hurdles to be overcome).

The technical advances were followed by theoretical progress aimed at understanding the solution's structure. The goal is to discover some of the solution's properties before actually computing it, and sometimes even without a complete solution. The theoretical analysis of PDEs is not merely of academic interest, but rather has many applications. It should be stressed that there exist very complex equations that cannot be solved even with the aid of supercomputers. All we can do in these cases is to attempt to obtain qualitative information on the solution. In addition, a deep important question relates to the formulation of the equation and its associated side conditions. In general, the equation originates from a model of a physical or engineering problem. It is not automatically obvious that the model is indeed consistent in the sense that it leads to a solvable PDE. Furthermore, it is desired in most cases that the solution will be unique, and that it will be stable under small perturbations of the data. A theoretical understanding of the equation enables us to check whether these conditions are satisfied. As we shall see in what follows, there are many ways to solve PDEs, each way applicable to a certain class of equations. Therefore it is important to have a thorough analysis of the equation before (or during) solving it.

Partial differential equations come in three types. For a function of two variables $T = T(x, y)$, the general second-order linear PDE has the form

$$a\frac{\partial^2 T(x,y)}{\partial x^2} + 2b\frac{\partial^2 T(x,y)}{\partial x \partial y} + c\frac{\partial^2 T(x,y)}{\partial y^2} + d\frac{\partial T(x,y)}{\partial x} + e\frac{\partial T(x,y)}{\partial y} + fT(x,y) = A \tag{C.84}$$

where a, b, c, d, e, f, g may depend on x and y only. We call Eq. (C.17)

$$\begin{aligned} &\textit{elliptic} && \text{if} \quad b^2 - ac < 0 \\ &\textit{hyperbolic} && \text{if} \quad b^2 - ac > 0 \\ &\textit{parabolic} && \text{if} \quad b^2 - ac = 0 \end{aligned}$$

Example C.1 : Assuming $T = T(x)$ as function of one variable

1. The heat equation

$$\frac{\partial T}{\partial t} = c\frac{\partial^2 T}{\partial x^2}$$

   is parabolic
2. The wave equation

$$\frac{\partial^2 T}{\partial t^2} = a^2\frac{\partial^2 T}{\partial x^2}$$

   is hyperbolic
3. The Laplace equation

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0$$

   is elliptic

C.7.2 Classification

We pointed out in the previous section that PDEs are often classified into different types. In fact, there exist several such classifications. Some of them will be described here.

- **The order of an equation**
  The first classification is according to the *order* of the equation. The order is defined to be the order of the highest derivative in the equation. If the highest derivative is of order k. Thus, for example, the equation $u_{tt} + u_{xxxx} = 0$ is called fourth-order equation.

- **Linear equations**
  Another classification is into two groups: linear versus nonlinear equations. An equation is called *linear* if in Eq. (C.83), F is a linear function of the unknown function u and its derivatives. Thus, for example, the equation $x^7 u_x + e^{xy} u_y + \sin(x^2 + y^2)u = x^3$ is a linear equation, while $u_x^2 + u_y^2 = 1$ is a nonlinear equation. The nonlinear equations are often further classified into subclasses according to the type of the nonlinearity. Generally speaking, the nonlinearity is more pronounced when it appears in a higher derivative. For example, the following two equations are both nonlinear:

$$u_{xx} + u_{yy} = u^3 \tag{C.85}$$

$$u_{xx} + u_{yy} = |\nabla u|^2 u \tag{C.86}$$

Here $|\nabla u|$ denotes the norm of the gradient of u While Eq. (C.86) is nonlinear, it is still linear as a function of the highest-order derivative. Such nonlinearity is called *quasilinear*. On the other hand in Eq. (C.85) the nonlinearity is only in the unknown function. Such equations are often called *semilinear*.

- **Scalar equations versus systems of equations**
  A single PDE with just one unknown function is called a *scalar equation*. In contrast, a set of m equations with l unknown functions is called a *system* of m equations.

C.7.3 Differential Operators and Superposition Principle

A function has to be k times differentiable in order to be a solution of an equation of order k. For this purpose we define the set $C^k(D)$ to be the set of all functions that are k times continuously differentiable in D. In particular, we denote the set of continuous functions in D by $C^0(D)$, or $C(D)$. A function in the set C^k that satisfies a PDE of order k, will be called a *classical* (or *strong*) solution of the PDE. Mappings between different function sets are called *operators*. The operation of an operator L on a function u will be denoted by $L[u]$. In particular, we shall deal in this book with operators defined by partial derivatives of functions. Such operators, which are in fact mappings between different C^k classes, are called *differential operators*.

An operator that satisfies a relation of the form

$$L[a_1u_1 + a_2u_2] = a_1L[u_1] + a_2L[u_2]$$

where a_1 and a_2 are arbitrary constant, and u_1, u_2 are arbitrary functions are called a *linear operator*. A linear differential equation naturally defines a linear operator: the equation can be expressed as $L[u] = f$, where L is a linear operator and f is a given function. A linear differential equation of the form $L[u] = 0$, where L is a linear operator, is called a *homogeneous equation*. For example, define the operator $L = \dfrac{\partial^2}{\partial x^2} - \dfrac{\partial^2}{\partial y^2}$. The equation

$$L[u] = u_{xx} - u_{yy} = 0$$

is homogeneous equation, while the equation

$$L[u] = u_{xx} - u_{yy} = x^2$$

is an example of *nonhomogeneous equation*.

C.8 Heat Equation

We begin by formulating the equations of heat flow describing the transfer of thermal energy. Heat energy is caused by the agitation of molecular matter. Two basic processes take place in order for thermal energy to move;

1. Conduction
   This results from the collision of neighboring molecules in which the kinetic energy of vibration of one molecule is transferred to its nearest neighbor. Thermal energy is thus spread by conduction even if the molecules themselves do not move their location appreciably.
2. Convection
   In addition to conduction, if a vibrating molecule moves from one region to another, it takes its thermal energy with it. This type of movement of thermal energy is called convection.
3. Radiation
   Radiation is energy that comes from a source and travels through some material or through space. Light, heat and sound are types of radiation. The kind of radiation discussed in this presentation is called ionizing radiation because it can produce charged particles (ions) in matter. Ionizing radiation is produced by unstable atoms. Unstable atoms differ from stable atoms because they have an excess of energy or mass or both.

Before we get into actually solving partial differential equations and before we even start discussing the method of separation of variables we want to spend a little bit of time talking about the two main partial differential equations that we'll be solving later on in the chapter. We'll look at the first one in this section and the second one in the next section. For those readers that have access to web site, I encourage them to go to the following link to go to and get more information on the Paul's Online Math Notes

http://tutorial.math.lamar.edu/

The first partial differential equation that we'll be looking at once we get started with solving will be the heat equation, which governs the temperature distribution in an object. We are going to give several forms of the heat equation for reference purposes, but we will only be really solving one of them.

We will start out by considering the temperature in a 1D bar of length L. What this means is that we are going to assume that the bar starts off at $x = 0$ and ends when we reach $x = L$. We are also going to so assume that at any location, x the temperature will be constant an every point in the cross section at that x. In other

words, temperature will only vary in x and we can hence consider the bar to be a 1D bar. Note that with this assumption the actual shape of the cross section (i.e., circular, rectangular, etc.) doesn't matter.

Note that the 1D assumption is actually not all that bad of an assumption as it might seem at first glance. If we assume that the lateral surface of the bar is perfectly insulated (i.e. no heat can flow through the lateral surface) then the only way heat can enter or leave the bar as at either end. This means that heat can only flow from left to right or right to left and thus creating a 1D temperature distribution.

The assumption of the lateral surfaces being perfectly insulated is of course impossible, but it is possible to put enough insulation on the lateral surfaces that there will be very little heat flow through them and so, at least for a time, we can consider the lateral surfaces to be perfectly insulated.

Okay, let's now get some definitions out of the way before we write down the first form of the heat equation.

$$u(x,t) = \text{Temperature at point } x \text{ and any time } t.$$

$$c(x) = \text{Specific Heat.}$$

$$\rho(x) = \text{Mass Density.}$$

$$\varphi(x,t) = \text{Heat Flux.}$$

$$Q(x,t) = \text{Heat energy generated per unit volume per unittime.}$$

We should probably make a couple of comments about some of these quantities before proceeding.

The specific heat, $c(x) > 0$, of a material is the amount of heat energy that it takes to raise one unit of mass of the material by one unit of temperature. As indicated we are going to assume, at least initially, that the specific heat may not be uniform throughout the bar. Note as well that in practice the specific heat depends upon the temperature. However, this will generally only be an issue for large temperature differences (which in turn depends on the material the bar is made out of) and so we're going to assume for the purposes of this discussion that the temperature differences are not large enough to affect our solution.

The mass density, $\rho(x)$, is the mass per unit volume of the material. As with the specific heat we're going to initially assume that the mass density may not be uniform throughout the bar.

The heat flux, $\varphi(x,t)$, is the amount of thermal energy that flows to the right per unit surface area per unit time. The "flows to the right" bit simply tells us that if $\rho(x) > 0$ for some x and t then the heat is flowing to the right at that point and time. Likewise if $\rho(x) < 0$ then the heat will be flowing to the left at that point and time.

The final quantity we defined above is $Q(x,t)$ and this is used to represent any external sources or sinks (i.e. heat energy taken out of the system) of heat energy. If $Q(x,t) > 0$ then heat energy is being added to the system at that location

and time and if $Q(x,t) < 0$ then heat energy is being removed from the system at that location and time.

With these quantities the heat equation is,

$$c(x)\rho(x)\frac{\partial u(x,t)}{\partial t} = -\frac{\partial \varphi(x,t)}{\partial x} + Q(x,t) \tag{C.87}$$

While this is a nice form of the heat equation it is not actually something we can solve. In this form there are two unknown functions, u and φ, and so we need to get rid of one of them. With Fourier's law we can easily remove the heat flux from this equation.

Fourier's law states that,

$$\varphi(x,t) = -K_0(x)\frac{\partial u(x,t)}{\partial t}$$

where $K_0(x) > 0$ is the thermal conductivity of the material and measures the ability of a given material to conduct heat. The better a material can conduct heat the larger $K_0(x)$ will be. As noted the thermal conductivity can vary with the location in the bar. Also, much like the specific heat the thermal conductivity can vary with temperature, but we will assume that the total temperature change is not so great that this will be an issue and so we will assume for the purposes here that the thermal conductivity will not vary with temperature.

Fourier's law does a very good job of modeling what we know to be true about heat flow. First, we know that if the temperature in a region is constant, i.e. $\frac{\partial u(x,t)}{\partial x} = 0$, then there is no heat flow.

Next, we know that if there is a temperature difference in a region we know the heat will flow from the hot portion to the cold portion of the region. For example, if it is hotter to the right then we know that the heat should flow to the left. When it is hotter to the right then we also know that $\frac{\partial u(x,t)}{\partial x} > 0$ (i.e., the temperature increases as we move to the right) and so we'll have $\varphi(x,t) < 0$ and so the heat will flow to the left as it should. Likewise, if $\frac{\partial u(x,t)}{\partial x} < 0$ (i.e., it is hotter to the left) then we'll have $\varphi(x,t) > 0$ and heat will flow to the right as it should.

Finally, the greater the temperature difference in a region (i.e. the larger $\frac{\partial u(x,t)}{\partial x}$ is) then the greater the heat flow.

So, if we plug Fourier's law into Eq. (C.87), we get the following form of the heat equation,

$$c(x)\rho(x)\frac{\partial u(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(K_0(x)\frac{\partial (u,x)}{\partial x}\right) + Q(x,t) \tag{C.88}$$

Note that we factored the minus sign out of the derivative to cancel against the minus sign that was already there. We cannot however, factor the thermal

conductivity out of the derivative since it is a function of x and the derivative is with respect to x.

Solving Eq. (C.88) is quite difficult due to the non uniform nature of the thermal properties and the mass density. So, let's now assume that these properties are all constant, i.e.,

$$c(x) = c \quad \rho(x) = \rho \quad K_0(x) = K_0$$

where c, ρ and K_0 are now all fixed quantities. In this case we generally say that the martial in the bar is **uniform**. Under these assumptions the heat equation becomes,

$$c\rho \frac{\partial u}{\partial t} = K_0 \frac{\partial^2 u}{\partial x^2} + Q(x,t) \tag{C.89}$$

For a final simplification to the heat equation let's divide both sides by $c\rho$ and define the thermal **diffusivity** to be,

$$k = \frac{K_0}{c\rho}$$

The heat equation is then,

$$\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2} + \frac{Q(x,t)}{c\rho} \tag{C.90}$$

To most people this is what they mean when they talk about the heat equation and in fact it will be the equation that we'll be solving. Well, actually we'll be solving Eq. (C.90) with no external sources, i.e. $Q(x,t) = 0$, but we'll be considering this form when we start discussing separation of variables in a couple of sections. We'll only drop the sources term when we actually start solving the heat equation.

Now that we've got the 1D heat equation taken care of we need to move into the initial and boundary conditions we'll also need in order to solve the problem. If you go back to any of our solutions of ordinary differential equations that we've done in previous sections you can see that the number of conditions required always matched the highest order of the derivative in the equation.

In partial differential equations the same idea holds except now we have to pay attention to the variable we're differentiating with respect to as well. So, for the heat equation we've got a first-order time derivative and so we'll need one initial condition and a second order spatial derivative and so we'll need two boundary conditions.

The initial condition that we'll use here is

$$u(x,0) = f(x)$$

and we don't really need to say much about it here other than to note that this just tells us what the initial temperature distribution in the bar is.

The boundary conditions will tell us something about what the temperature and/or heat flow is doing at the boundaries of the bar. There are four of them that are fairly common boundary conditions.

The first type of boundary conditions that we can have would be the prescribed temperature boundary conditions, also called **Dirichlet conditions**. The prescribed temperature boundary conditions are,

$$u(0,t) = g_1(x) \quad u(L,t) = g_2(t)$$

The next type of boundary conditions are **prescribed heat flux**, also called **Neumann conditions**. Using Fourier's law these can be written as,

$$-K_0(0)\frac{\partial u(0,t)}{\partial x} = \varphi_1(t) \quad -K_0(L)\frac{\partial u(L,t)}{\partial x} = \varphi_2(t)$$

If either of the boundaries is **perfectly insulated**, i.e., there is no heat flow out of them then these boundary conditions reduce to,

$$\frac{\partial u(0,t)}{\partial x} = 0 \quad \frac{\partial u(L,t)}{\partial x} = 0$$

and note that we will often just call these particular boundary conditions insulated boundaries and drop the "perfectly" part.

The third type of boundary conditions use **Newton's law of cooling** and are sometimes called **Robins conditions**. These are usually used when the bar is in a moving fluid and note we can consider air to be a fluid for this purpose.

Here are the equations for this kind of boundary condition.

$$-K_0(0)\frac{\partial u(0,t)}{\partial x} = H[u(0,t) - g_1(t)] \quad -K_0(L)\frac{\partial u(L,t)}{\partial x} = H[u(L,t) - g_2(t)]$$

where H is a positive quantity that is experimentally determined $g_1(t)$ and $g_2(t)$ give the temperature of the surrounding fluid at the respective boundaries.

Note that the two conditions do vary slightly depending on which boundary we are at. At $x = 0$ we have a minus sign on the right side while we don't at $x = L$. To see why this is let's first assume that at $x = 0$ we have $u(0,t) > g_1(t)$. In other words the bar is hotter than the surrounding fluid and so at $x = 0$ the heat flow (as given by the left side of the equation) must be to the left, or negative since the heat will flow from the hotter bar into the cooler surrounding liquid. If the heat flow is negative then we need to have a minus sign on the right side of the equation to make sure that it has the proper sign.

If the bar is cooler than the surrounding fluid at $x = 0$, i.e. $u(0,t) < g_1(t)$ we can make a similar argument to justify the minus sign. We'll leave it to you to verify this.

If we now look at the other end, $x = L$, and again assume that the bar is hotter than the surrounding fluid or, $u(L,t) > g_2(t)$. In this case the heat flow must be to the right, or be positive, and so in this case we can't have a minus sign. Finally, we'll again leave it to you to verify that we can't have the minus sign at $x = L$ is the bar is cooler than the surrounding fluid as well.

Note that we are not actually going to be looking at any of these kinds of boundary conditions here. These types of boundary conditions tend to lead to boundary value problems such as Example 5 in the Eigenvalues and Eigenfunctions Sect. 3 of this appendix. As we saw in that example it is often very difficult to get our hands on the eigenvalues and as we'll eventually see we will need them.

It is important to note at this point that we can also mix and match these boundary conditions so to speak. There is nothing wrong with having a prescribed temperature at one boundary a prescribed flux at the other boundary for example so don't always expect the same boundary condition to show up at both ends. This warning is more important that it might seem at this point because once we get into solving the heat equation we are going to have the same kind of condition on each end to simplify the problem somewhat.

The final type of boundary conditions that we'll need here are periodic boundary conditions. Periodic boundary conditions are,

$$u(-L,t) = u(L,t) \quad \frac{\partial u}{\partial x}(-L,t) = \frac{\partial}{\partial x}u(L,t)$$

Note that for these kinds of boundary conditions the left boundary tends to be $x = -L$ instead of $x = 0$ as we were using in the previous types of boundary conditions. The periodic boundary conditions will arise very naturally from a couple of particular geometries that we'll be looking at down the road.

We will now close out this section with a quick look at the 2D and 3D version of the heat equation. However, before we jump into that we need to introduce a little bit of notation first.

The **del** operator is defined to be,

$$\nabla = \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j}\nabla = \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k}$$

depending on whether we are in 2 or 3 dimensions. Think of the del operator as a function that takes functions as arguments (instead of numbers as we're used to). Whatever function we "plug" into the operator gets put into the partial derivatives.

So, for example in 3D we would have

$$\nabla f = \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k}$$

This of course is also the gradient of the function $f(x, y, z)$.

The **del** operator also allows us to quickly write down the divergence of a function. So, again using 3D as an example the divergence of $f(x, y, z)$ can be written as the dot product of the **del** operator and the function. Or,

$$\nabla \cdot f = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}$$

Finally, we will also see the following show up in the our work,

$$\nabla \cdot (\nabla f) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial z}\right)$$

This is usually denoted as,

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

and is called the Laplacian. The 2-D version of course simply doesn't have the third term.

Okay, we can now into the 2D and 3D version of the heat equation and where ever the del operator and or Laplacian appears assume that it is the appropriate dimensional version.

The higher dimensional version of Eq. (C.83) is,

$$c\rho \frac{\partial u}{\partial t} = -\nabla \cdot \varphi + Q \tag{C.91}$$

and note that the specific heat, c, and mass density, ρ, are may not be uniform and so may be functions of the spatial variables. Likewise, the external sources term, Q, may also be a function of both the spatial variables and time.

Next, the higher dimensional version of Fourier's law is,

$$\varphi = -K_0 \nabla u$$

where the thermal conductivity, K_0, is again assumed to be a function of the spatial variables.

If we plug this into Eq. (C.91) we get the heat equation for a non uniform bar (i.e., the thermal properties may be functions of the spatial variables) with external sources/sinks,

$$c\rho\frac{\partial u}{\partial t} = \nabla\cdot(K_0\nabla u) + Q \tag{C.92}$$

If we now assume that the specific heat, mass density and thermal conductivity are constant (i.e., the bar is uniform) the heat equation becomes,

$$\frac{\partial u}{\partial t} = k\nabla^2 u + \frac{Q}{c\rho} \tag{C.93}$$

where we divided both sides by $c\rho$ to get the thermal diffusivity, k in front of the Laplacian.

The initial condition for the 2-D or 3-D heat equation is,

$$u(x,y,t) = f(x,y) \quad \text{or} \quad u(x,y,z,t) = f(x,y,z)$$

depending upon the dimension we're in.

The prescribed temperature boundary condition becomes,

$$u(x,y,t) = T(x,y,t) \quad \text{or} \quad u(x,y,z,t) = T(x,y,z,t)$$

where (x,y) or (x,y,z), depending upon the dimension we're in, will range over the portion of the boundary in which we are prescribing the temperature.

The prescribed heat flux condition becomes,

$$-K_0\nabla u\cdot\vec{n} = \varphi(t)$$

where the left side is only being evaluated at points along the boundary and is the outward unit normal on the surface.

Newton's law of cooling will become,

$$-K_0\nabla u\cdot\vec{n} = H(u - u_B)$$

where H is a positive quantity that is experimentally determine, u_B is the temperature of the fluid at the boundary and again it is assumed that this is only being evaluated at points along the boundary.

We don't have periodic boundary conditions here as they will only arise from specific 1D geometries.

We should probably also acknowledge at this point that we'll not actually be solving Eq. (C.93) at any point, but we will be solving a special case of it in the Laplace's Equation Sect. 5.1 below.

For those readers that have access to website, I encourage you to go to the following link and to get more information on the Paul's Online Math Notes

http://tutorial.math.lamar.edu/

C.8.1 Laplace's Equation

The next partial differential equation that we're going to solve is the 2D Laplace's equation,

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

A natural question to ask before we start learning how to solve this is does this equation come up naturally anywhere? The answer is a very resounding yes! If we consider the 2-D heat equation,

$$\frac{\partial u}{\partial t} = k\nabla^2 u + \frac{Q}{c\rho}$$

We can see that Laplace's equation would correspond to finding the equilibrium solution (i.e. time independent solution) if there were not sources. So, this is an equation that can arise from physical situations.

How we solve Laplace's equation will depend upon the geometry of the 2-D object we're solving it on. Let's start out by solving it on the rectangle given by $0 \le x \le L$, $0 \le y \le H$. For this geometry Laplace's equation along with the four boundary conditions will be,

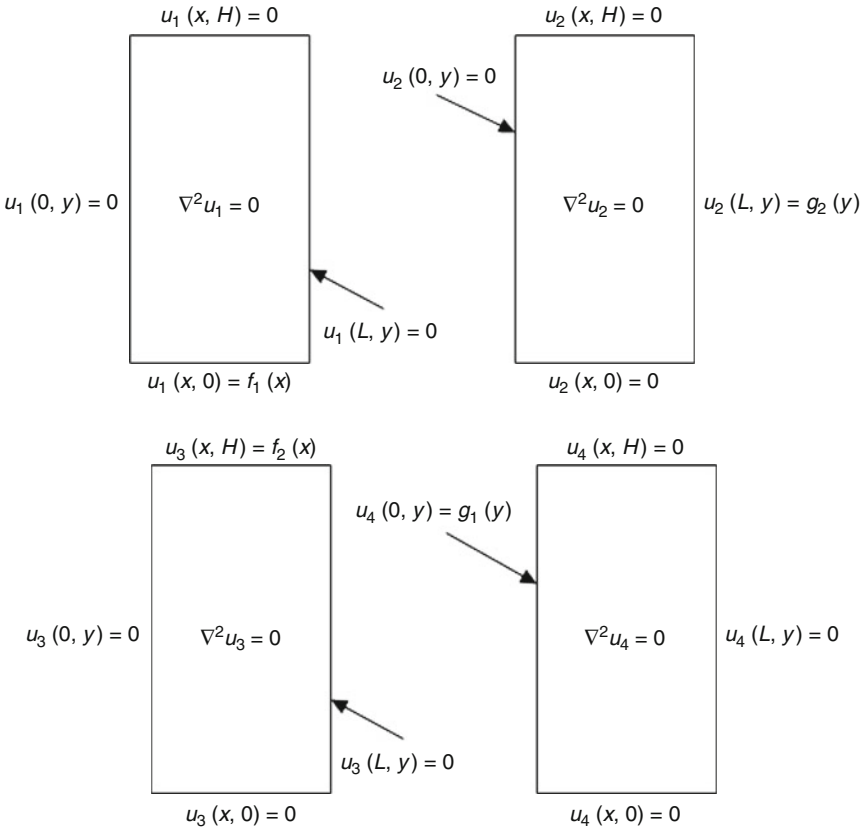
$$\begin{aligned} &\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \\ &u(0,y) = g_1(y) \qquad\qquad u(L,y) = g_2(y) \\ &u(x,0) = f_1(x) \qquad\qquad u(x,H) = f_2(y) \end{aligned} \tag{C.94}$$

One of the important things to note here is that unlike the heat equation we will not have any initial conditions here. Both variables are spatial variables and each variable occurs in a 2nd order derivative and so we'll need two boundary conditions for each variable.

Next, let's notice that while the partial differential equation is both linear and homogeneous the boundary conditions are only linear and are not homogeneous. This creates a problem because separation of variables requires homogeneous boundary conditions.

To completely solve Laplace's equation we're in fact going to have to solve it four times. Each time we solve it only one of the four boundary conditions can be nonhomogeneous while the remaining three will be homogeneous.

The four problems are probably best shown with a quick sketch so let's consider the following sketch.


$u_1(x, H) = 0$
$u_2(x, H) = 0$
$u_2(0, y) = 0$
$u_1(0, y) = 0$
$\nabla^2 u_1 = 0$
$\nabla^2 u_2 = 0$
$u_2(L, y) = g_2(y)$
$u_1(L, y) = 0$
$u_1(x, 0) = f_1(x)$
$u_2(x, 0) = 0$
$u_3(x, H) = f_2(x)$
$u_4(x, H) = 0$
$u_4(0, y) = g_1(y)$
$u_3(0, y) = 0$
$\nabla^2 u_3 = 0$
$\nabla^2 u_4 = 0$
$u_4(L, y) = 0$
$u_3(L, y) = 0$
$u_3(x, 0) = 0$
$u_4(x, 0) = 0$

Now, once we solve all four of these problems the solution to our original system, (Eq. C.94), will be,

$$u(x,y) = u_1(x,y) + u_2(x,y) + u_3(x,y) + u_4(x,y)$$

Because we know that Laplace's equation is linear and homogeneous and each of the pieces is a solution to Laplace's equation then the sum will also be a solution. Also, this will satisfy each of the four original boundary conditions. We'll verify the first one and leave the rest to you to verify.

$$u(x,0) = u_1(x,0) + u_2(x,0) + u_3(x,0) + u_4(x,0) = f_1(x) + 0 + 0 + 0 = f_1(x)$$

In each of these cases the lone nonhomogeneous boundary condition will take the place of the initial condition in the heat equation problems that we solved a couple of sections ago. We will apply separation of variables to the each problem and find a product solution that will satisfy the differential equation and the three homogeneous boundary conditions. Using the Principle of Superposition we'll find a solution to the problem and then apply the final boundary condition to determine

the value of the constant(s) that are left in the problem. The process is nearly identical in many ways to what we did when we were solving the heat equation.

We're going to do two of the cases here and we'll leave the remaining two for you to do.

Example 1: Find a solution to the following partial differential equation.

$$\nabla^2 u_4 = \frac{\partial^2 u_4}{\partial x^2} + \frac{\partial^2 u_4}{\partial y^2} = 0$$

$$u_4(0,y) = g_1(y) \qquad u_4(L,y) = 0$$

$$u_4(x,0) = 0 \qquad u_4(x,H) = 0$$

Solution: We'll start by assuming that our solution will be in the form of the following using separation of variable method,

$$u_4(x,y) = h(x)\varphi(y)$$

and then recall that we performed separation of variables on this problem (with a small change in notation) back in Example 5 of the Separation of Variables Sect. 2.1, Appendix I. So from that problem we know that separation of variables yields the following two ordinary differential equations that we'll need to solve.

$$\frac{d^2h}{dx^2} - \lambda h = 0 \quad \frac{d^2\varphi}{dy^2} + \lambda\varphi = 0$$

$$h(L) = 0 \qquad \varphi(0) = 0 \qquad \varphi(H) = 0$$

Note that in this case, unlike the heat equation we must solve the boundary value problem first. Without knowing what λ is there is no way that we can solve the first differential equation here with only one boundary condition since the sign of λ will affect the solution.

Let's also notice that we solved the boundary value problem in Example 1 of Solving the Heat Equation and so there is no reason to resolve it here. Taking a change of letters into account the eigenvalues and eigenfunctions for the boundary value problem here are

$$\lambda_n = \left(\frac{n\pi}{H}\right)^2 \quad \varphi_n(y) = \sin\left(\frac{n\pi}{H}\right) \quad n = 1,2,3,\cdots$$

Now that we know what the eigenvalues are let's write down the first differential equation with λ plugged in.

Because the coefficient of $h(x)$ in the differential equation above is positive we know that a solution to this is,

$$h(x) = c_1 \cosh\left(\frac{n\pi}{H}\right) + c_2 \sinh\left(\frac{n\pi}{H}\right)$$

However, this is not really suited for dealing with the $h(L) = 0$ boundary condition. So, let's also notice that the following is also a solution.

$$h(x) = c_1 \cosh\left(\frac{n\pi(x-L)}{H}\right) + c_2 \sinh\left(\frac{n\pi(x-L)}{H}\right)$$

You should verify this by plugging this into the differential equation and checking that it is in fact a solution. Applying the lone boundary condition to this "shifted" solution gives,

$$0 = h(L) = c_1$$

The solution to the first differential equation is now,

$$h(x) = c_2 \sin h\left(\frac{n\pi(x-L)}{H}\right)$$

and this is all the farther we can go with this because we only had a single boundary condition. That is not really a problem however because we now have enough information to form the product solution for this partial differential equation. A product solution for this partial differential equation is,

$$u_n(x,y) = B_n \sin h\left(\frac{n\pi(x-L)}{H}\right) \sin\left(\frac{n\pi}{H}\right) \quad n = 1, 2, 3, \cdots$$

The Principle of Superposition then tells us that a solution to the partial differential equation is,

$$u_4(x,y) = \sum_{n=1}^{\infty} B_n \sin h\left(\frac{n\pi(x-L)}{H}\right) \sin\left(\frac{n\pi}{H}\right)$$

and this solution will satisfy the three homogeneous boundary conditions. To determine the constants all we need to do is apply the final boundary condition.

$$u_4(0,y) = g_1(y) = \sum_{n=1}^{\infty} B_n \sin h\left(\frac{n\pi(-L)}{H}\right) \sin\left(\frac{n\pi y}{H}\right)$$

Now, in the previous problems we've done this has clearly been a Fourier series of some kind and in fact it still is. The difference here is that the coefficients of the Fourier sine series are now,

$$B_n \sin h\left(\frac{n\pi(-L)}{H}\right)$$

instead of just B_n. We might be a little more tempted to use the orthogonality of the sines to derive formulas for the B_n, however we can still reuse the work that we've done previously to get formulas for the coefficients here.

Remember that a Fourier sine series is just a series of coefficients (depending on n) times a sine. We still have that here, except the "coefficients" are a little messier this time that what we saw when we first dealt with Fourier series. So, the coefficients can be found using exactly the same formula from the **Fourier Sine Series** section of a function on $0 \le y \le H$ we just need to be careful with the coefficients.

$$B_n \sin h\left(\frac{n\pi(-L)}{H}\right) = \frac{2}{H}\int_0^H g_1(y) \sin\left(\frac{n\pi y}{H}\right) dy \quad n = 1, 2, 3, \cdots$$

$$B_n \frac{2}{H \sin h\left(\frac{n\pi(-L)}{H}\right)} \int_0^H g_1(y) \sin\left(\frac{n\pi y}{H}\right) dy \qquad n = 1, 2, 3, \cdots$$

The formulas for the B_n are a little messy this time in comparison to the other problems we've done but they aren't really all that messy.

For more complex examples and further details we urge reader to the following site http://tutorial.math.lamar.edu/Classes/DE/LaplacesEqn.aspx

C.9 References

1. Richard Haberman. (1983) Elementary applied partial differential equations with fourier series and boundary value problems. 2nd Edition, Prentice-Hall, Inc
2. Erich Zauderer (1989) Partial differential equations of applied mathematics. 2nd Edition, John Wiley
3. Yehuda Pinchover and Jacob Rubinstein (2005) An introduction to partial differential equations. Cambridge University Press
4. Dennis G. Zill. A first course in differential equations with modeling applications. 7th Edition, Brooks/Cole Thomson Learning
5. Boyce We and Diprima Rc (2001) Elementary differential equations and boundary value problems. 7th Ed, John Wiley

Appendix D
Short Course in Complex Variables

In the study of Laser Physics and its interaction with materials as well as solving heat transfer and related physics of optics, a great saving in complexity of notation may be accomplished by using the notation of Complex Variables and Applications. The purpose of this section is to give a brief but self-contained exposition of basic complex variables analysis to bring to the forefront of physical ideas involved in solving heat transfer equations and provide knowledge of the field that is required for treatment of partial differential equations that results in solving such complex boundary value problems. The primary objective of this section is to develop those parts of the theory that are prominent in the applications of the subject. Note that the basic and principal analysis of complex variable numbers section (Sects. 1.1 to 1.10) at the beginning of this chapter has been borrowed from book by Brown and Churchill [1] and the rest has been gathered from different sections of Schaum's Outlines Complex Variables (i.e., courtesy of published) with an Introduction to Conformal Mapping and Its Application by Murray R. Spiegel [2] but it is formatted to the need of this book where we need to solve boundary value problems of heat diffusion and it use it in understanding of Physics of Laser Beam Interaction with materials. Also various different references (See references that have been mentioned at the end of this chapter) are used for different examples and problems presented in this Appendix.

D.1 Complex Numbers

A complex number is a number consisting of a real and imaginary part. It can be written in the form $a + ib$, where a and b are real numbers, and i is the standard imaginary unit with the property $i^2 = -1$ The complex numbers contain the ordinary real numbers, but extend them by adding in extra numbers and correspondingly expanding the understanding of addition and multiplication (Fig. D.1).

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

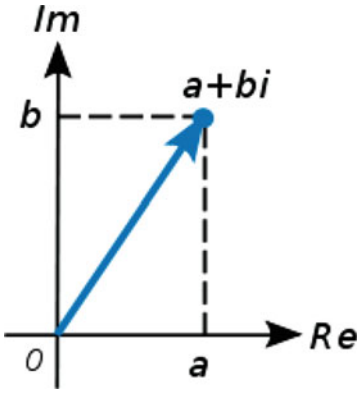


Fig. D.1 A complex number can be visually represented as a pair of numbers forming a vector on a diagram called an Argand diagram, representing the complex plane

If $z = a + ib$, the real part a is denoted $Re(z)$, and the imaginary part b is denoted $Im(z)$. The complex numbers (C) are regarded as an extension of the real numbers (R) by considering every real number as a complex number with an imaginary part of zero. The real number a is identified with the complex number $a + 0i$. Complex numbers with a real part of zero (Re(z)=0) are called imaginary numbers. Instead of writing $0 + ib$, that imaginary number is usually denoted as just ib. If b equals 1, instead of using $0 + 1i$ or $1i$, the number is denoted as i.

In some disciplines (in particular, electrical engineering, where i is a symbol for current), the imaginary unit i is instead written as j, so complex numbers are sometimes written as $a + jb$ or $a + bj$.

D.2 Branch Points and Branch Lines

Suppose that we are given the function $w = z^{1/2}$. Suppose further that we allow z to make a complete circuit (counterclockwise) around the origin starting from point A (See Fig. D.2). We have $z = re^{i\theta}$, $w = \sqrt{r}e^{i\theta/2}$ so that at A, $\theta = \theta_1$ and $w = \sqrt{r}e^{i\theta_1/2}$. After a complete circuit back to A, $\theta = \theta_1 + 2\pi$ and $w = \sqrt{r}e^{i(\theta_1+2\pi)/2} = -\sqrt{r}e^{i\theta_1/2}$. Thus we have not attached the same value of w with which we started. However, by making a second complete circle back to A, i.e. $\theta = \theta_1 + 4\pi$, $w = \sqrt{r}e^{i(\theta_1+4\pi)/2} = \sqrt{r}e^{i\theta_1/2}$ and we then do obtain the same value of w with which we started. We can describe the above by stating that if $0 \le \theta < 2\pi$ we are on one branch of the multiple-valued function $z^{1/2}$, while if $2\pi \le \theta < 4\pi$ we are on the other branch of the function.

It is clear that each branch of the function is single-valued. In order to keep the function single-valued, we set up an artificial barrier such as OB where B is at infinity [although any other line from O can be used] which we agree not to cross. This barrier [drawn heavy in the figure] is called a *branch cut*, and point O is called

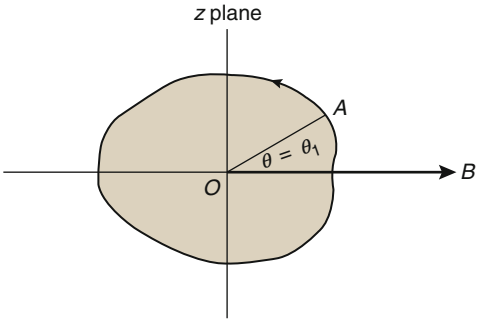


Fig. D.2 Branch Points and Branch Lines

a *branch point*. It should be noted that a circle around any point other than $z = 0$ does not lead to different values; thus $z = 0$ is the only finite branch point.

D.3 Singular Points

A point at which $f(z)$ fails to be analytic is called a *singular point* or *singularity* of $f(z)$. Various types of singularities exist.

D.3.1 Isolated Singularities

The point $z = z_0$ is called an *isolated singularity* or *isolated singular point* of $f(z)$ if we can find $\delta > 0$ such that the circle $|z - z_0| = \delta$ encloses no singular point other than z_0 (i.e., there exists a deleted δ neighborhood of z_0 containing no singularity). If no such δ can be found, we call z_0 a *non-isolated singularity*.

If z_0 is not a singular point and we can find $\delta > 0$ such that $|z - z_0| = \delta$ encloses no singular point, then we call z_0 an *ordinary point* of $f(z)$.

D.3.2 Poles

If we can find a positive integer n such that $\lim_{z \to z_0} (z - z_0)^n f(z) = A \neq 0$, then $z = z_0$ is called a *pole of order n*. If $n = 1$, z_0 is called a *simple pole*.

Example 1: $f(z) = \frac{1}{(z-2)^3}$ has a pole of order 3 at $z = 2$

Example 2: $f(z) = \frac{3z-2}{(z-1)^2(z+1)(z-4)}$ has a pole of order 2 at $z = 1$, and simple poles at $z = -1$ and $z = 4$.

If $g(z) = (z - z_0)^n f(z)$, where $f(z_0) \neq 0$ and n is a positive integer, then $z = z_0$ is called a *zero of order n* of $g(z)$. If $n = 1$, z_0 is called a *simple zero*. In such case z_0 is a pole of order n of the function $1/g(z)$.

D.3.3 Branch Points

Branch points of multiple-valued functions, already considered in Sect. 2, are singular points.

Example 1: $f(z) = (z - 3)^{1/2}$ has a branch point at $z = 3$.

Example 2: $f(z) = \ln(z^2 + z - 2)$ has branch points where $z^2 + z - 2 = 0$, i.e. at $z = 1$ and $z = -2$

D.3.4 Removable Singularities

The singular point z_0 is called a *removable singularity* of $f(z)$ if $\lim_{z \to z_0} f(z)$ exists.

Example 1: The singular point $z = 0$ is a removable singularity of $f(z) = \frac{\sin z}{z}$ since $\lim_{z \to 0} \frac{\sin z}{z} = 1$

D.3.5 Essential Singularities

A singularity which is not a pole, branch point or removable singularity is called an *essential singularity*.

Example 1: $f(z) = e^{1/(z-2)}$ has an essential singularity at $z = 2$.

If a function is single-valued and has a singularity, then the singularity is either a pole or an essential singularity. For this reason a pole is sometimes called a *non-essential singularity*. Equivalently, $z = z_0$ is an essential singularity if we cannot find any positive integer n such that $\lim_{z \to z_0} (z - z_0)^n f(z) = A \neq 0$.

D.3.6 Singularities at Infinity

The type of singularity of $f(z)$ at $z = \infty$ [the point at infinity see Sect. 12.0] is the same as that of $f(1/w)$ at $w = 0$

Example 1: The function $f(z) = z^3$ has a pole of order 3 at $z = \infty$, since $f(1/w) = 1/w^3$ has a pole of order 3 at . $w = 0$

For methods of classifying singularities using infinite series (i.e., See Chap. 6 of reference [3] of this appendix)

D.4 Orthogonal Families

In this $w = f(z) = u(x, y) + iv(x, y)$ is analytic, then the one-parameter families of curves

$$u(x, y) = \alpha \quad v(x, y) = \beta \tag{D.1}$$

where α and β are constants, are *orthogonal* , i.e., each member of one family (shown heavy in Fig. D.3) is perpendicular to each member of the other family (shown dashed in Fig. D.4) at the point of intersection.

The corresponding image curves in the plane consisting of lines parallel to the u axes also from orthogonal families (See Fig. D.4).

D.5 Curves

If $\phi(t)$ and $\psi(t)$ are real functions of the real variable t assumed continuous in $t_1 \leq t \leq t_2$, the parametric equations

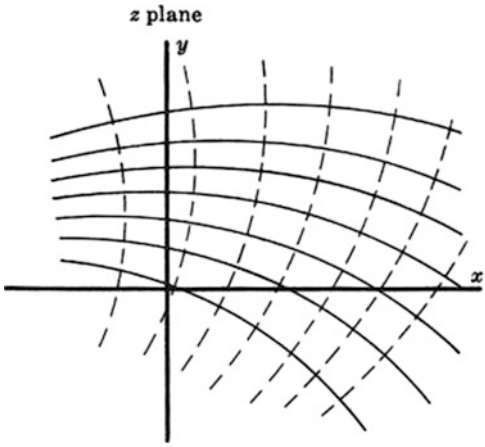


Fig. D.3 Orthogonal Families

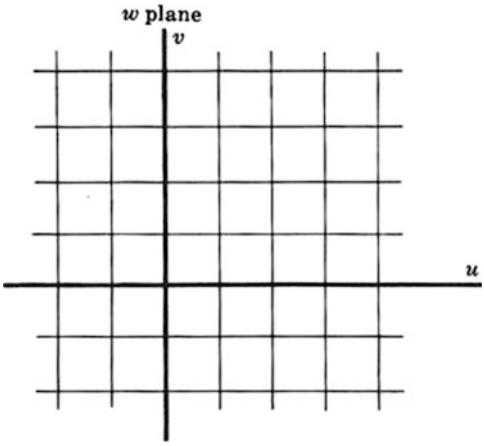


Fig. D.4 Orthogonal Families

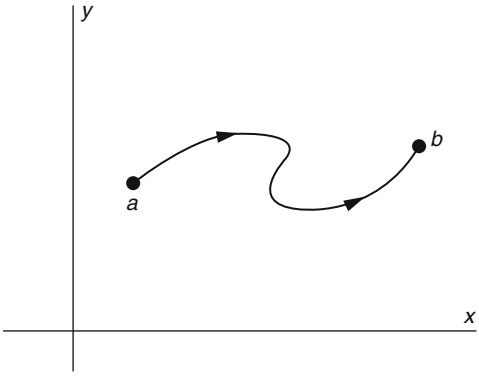


Fig. D.5 Curve Contour

$$z = x + iy = \phi(t) + i\psi(t) = z(t) \quad t_1 \leq t \leq t_2 \tag{D.2}$$

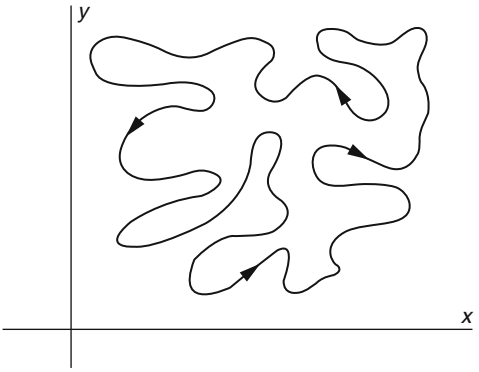
define a *continuous curve* or *arc* in the z plane joining points $a = z(t_1)$ and $b = z(t_2)$ (See Fig. D.5) below.

If $t_1 \neq t_2$ while $z(t_1) = z(t_2)$. i.e. $a = b$, the endpoints coincide and the curve is said to be *closed*. A closed curve which does not intersect itself anywhere is called a *simple closed curve*. For example the curve of Fig. D.5 is a simple closed curve while that of Fig. D.6 is not.

If $\phi(t)$ and $\psi(t)$ [and thus $z(t)$ have continuous derivative in $t_1 \leq t \leq t_2$ the curve is often called a *smooth curve* or *arc*. A curve which is composed of finite number of smooth arcs is called a piecewise or *sectionally smooth curve* or sometimes a *contour*. For example, the boundary of a square is a piecewise smooth curve or contour.

Unless otherwise specified, whenever we refer to a curve or simple closed curve we shall assume it to be piecewise smooth.

Fig. D.6 Curve Contour



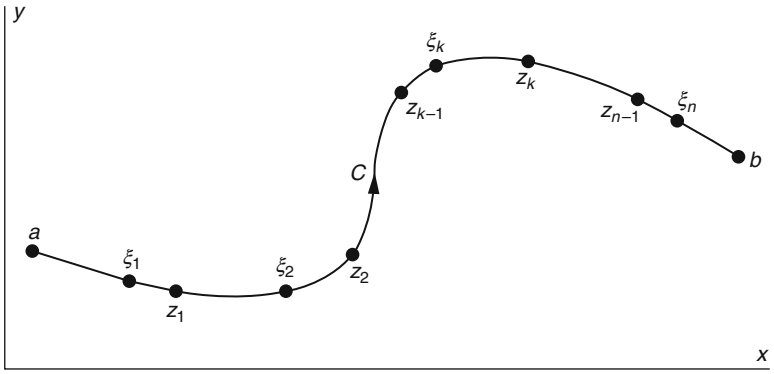


Fig. D.7 Curve C

D.6 Complex Integration and Cauchy's Theorem

In this section we introduce Cauchy's theorem and complex integration. Here is perhaps the most remarkable fact about analytic functions: the *values of an analytic function* $f(z)$ *on a closed loop* C dictate *its values at every point inside*. If z_0 is a point inside C, then the equation relating $f(z_0)$ to the known values of $f(z)$ on C is called the Cauchy integral formula.

D.6.1 Complex Line Integrals

Let $f(z)$ at all points of a curve C (See Fig. D.7) which we shall assume has a finite length, i.e. C is a *rectifiable curve*.

Subdivide C into n parts by means of points $z_1, z_2, \ldots, z_{n-1}$, chosen arbitrarily, and call $a = z_0, b = z_n$. On each are joining z_{k-1} to z_k [where $k = 1, 2, \ldots n$] choose a point ξ_k. From the sum

$$S_n = f(\xi_1)(z_1 - a) + f(\xi_2)(z_2 - z_1) + \ldots + f(\xi_n)(b - z_{n-1}) \tag{D.3}$$

On writing $z_k - z_{k-1} = \Delta z_k$, this becomes

$$S_n = \sum_{k=1}^{n} f(\xi_k)(z_k - z_{k-1}) = \sum_{k=1}^{n} f(\xi_k)\Delta z_k \tag{D.4}$$

Let the number of subdivision n increase in such a way that the largest of the chord lengths $|\Delta z_k|$ approaches zero. Then the sum S_n approaches a limit which does not depend on the mode of subdivision and we denote this limit by:

$$\int_a^b f(z)dz \quad \text{or} \quad \int_C f(z)dz \tag{D.5}$$

Called the *complex line integral* or briefly *line integral* of $f(z)$ along curve C, or the *definite integral* of $f(z)$ from a to b along curve C. In such case $f(z)$ is said to be *integrable* along C. Note that if $f(z)$ is analytic at all points of a region $\mathfrak{R}$ and if C is a curve lying in $\mathfrak{R}$ then $f(z)$ is certainly integrable along C.

D.6.2 Real Line Integrals

If $P(x, y)$ and $Q(x, y)$ are real functions of x and y continuous at all points of curve C, the *real line integral* of $Pdx + Qdy$ along curve C can be defined in a manner similar to that given above and is denoted by

$$\int_C [P(x, y)dx + Q(x, y)dy] \quad \text{or} \quad \int_C Pdx + Qdy \tag{D.6}$$

the second notation being used for brevity. If C is smooth and has parametric equations $x = \phi(t)$ and $y = \psi(t)$ where $t_1 \leq t \leq t_2$, the value of Eq. D.6 is given by

$$\int_{t_1}^{t_2} \left[P\{\phi(t), \psi(t)\}\phi'(t)dt + Q\{\phi(t), \psi(t)\}\psi'(t)dt\right]$$

Suitable modification can be made if C is piecewise smooth. See below example;

Example: Evaluate $\int_{(0,3)}^{(2,4)} \left(2y + x^2\right)dx + 3(3x - y)dy$ along

(a) The parabola $x = 2t$, $y = t^2 + 3$
(b) Straight lines from (0,3) to (2,3) and then from (2,3) to (2,4)
(c) A straight line from (0,3) to (2,4)

Solution:

(a) The point (0,3) and (2,4) on the parabola correspond to $t = 0$ and $t = 1$ respectively. Then the given integral equals

$$\int_{t=0}^{1} \left\{2\left(t^2 + 3\right) + (2t)^2\right\}dt + \left\{3(2t) - \left(t^2 + 3\right)\right\}2tdt$$

$$= \int_{t=0}^{1} \left(24t^2 + 12 - 2t^3 - 6t\right)dt = 33/2$$

(b) Along the straight line from (0,3) to (2,3), $y = 3$, $dy = 0$ and the line integral equals

$$\int_{x=0}^{2} \left(6 + x^2\right)dx + (3x - 3)0 = \int_{x=0}^{2} \left(6 + x^2\right)dx = 44/3$$

Along the straight line from (2,3) to (2,4), $x = 2$, $dx = 0$ and the line integral equals

$$\int_{y=3}^{4} (2y + 4)0 + (6 - y)dy = \int_{y=3}^{2} \left(6 + x^2\right)dx = 5/2$$

Then the required value = 44/3+5/2=103/6.

(c) An equation for the line joining (0,3) and (2,4) is $2y - x = 6$. Solving for x, we $y = \frac{1}{2}(x + 6)$ have $x = 2y - 6$. Then the line integral equals

$$\int_{y=3}^{4} \left\{2y + (2y - 6)^2\right\}2dy + \{3(2y - 6) - y\}dy$$

$$= \int_{3}^{4} \left(8y^2 - 39y + 54\right)dy = 97/6$$

The result can also be obtained by using $y = \frac{1}{2}(x + 6)$.

D.6.3 Connection Between Real and Complex Line Integrals

If $f(z) = u(x,y) + iv(x,y) = u + iv$ the complex line integral Eq. (D.94) can be expressed in terms of real line integrals as

$$\int_C f(z)dz = \int_C (u + iv)(dx + idy) = \int_C udx - vdu + i\int_C vdx + udv \qquad (D.7)$$

For this reason the above equation sometimes is taken as a definition of a complex line integral.

D.6.4 Properties of Integrals

If $f(z)$ and $g(z)$ are integrable along C, then

1. $\int_C \{f(z) + g(z)\}dz = \int_C f(z)dz + \int_C g(z)dz$
2. $\int_C Af(z)dz = A\int_C f(z)dz$ where $A =$ any constant.
3. $\int_C f(z)dz = -\int_C f(z)dz$
4. $\int_C f(z)dz = \int_a^m f(z)dz + \int_m^b f(z)dz$ where points a, b, m are on C.
5. $\left|\int_C f(z)dz\right| \leq ML$

   where $|f(z)| \leq ML$, i.e., is an *upper bound* of $|f(z)|$ on C, and L is the *length* of C.

There are various other ways in which the above properties can be described. For example if T and U and V are successive points on a curve, Equation D.6 can be then written as:

$$\int_{TUV} f(z)dz = -\int_{VUT} f(z)dz.$$

Similarly if C, C_1 and C_2 represent curves from a to b, a to m, and m to b, respectively, it is natural for us to consider $C = C_1 + C_2$ and to write Eq. D.7 as

$$\int_{C_1+C2} f(z)dz = \int_{C_1} f(z)dz + \int_{C_2} f(z)dz$$

D.6.5 Change of Variables

Let $z = g(\xi)$ be a continuous function of a complex variable $\xi = u + iv$. Suppose that curve C in the z plane corresponding to curve C' in the ζ plane and that the derivative $g'(\zeta)$ is continuous on C'. Then

$$\int_C f(z)dz = \int_{C'} f\{g(\zeta)\}g'(\zeta)dg \tag{D.8}$$

These conditions are certainly satisfied if g is analytic in a region containing curve C'.

D.6.6 Simply and Multiply Connected Regions

A region $\mathfrak{R}$ is called *simply-connected* if any simple closed curve see Sect. 18 which lies in $\mathfrak{R}$ can be shrunk to a point without leaving $\mathfrak{R}$. A region $\mathfrak{R}$ which is not simply-connected is called *multiply-connected.*

For example suppose $\mathfrak{R}$ is the region defined by $|z| < 2$ shown shaded in Fig. D.8a. If Γ is any simple closed curve lying in $\mathfrak{R}$ (i.e., whose points are in $\mathfrak{R}$), we see that it can be shrunk to a point which lies in $\mathfrak{R}$, and thus does not leave $\mathfrak{R}$, so that $\mathfrak{R}$ is simply-connected. On the other hand if $\mathfrak{R}$ is the region defined by $1 < |z| < 2$ shown shaded in Fig. D.8b, then there is a simple closed curve Γ lying in $\mathfrak{R}$, which cannot possibly be shrunk to a point without leaving $\mathfrak{R}$, so that $\mathfrak{R}$, is multiply-connected.

Intuitively, a simply-connected region is one which does not have any "holes" in it, while a multiply-connected region is one which does. Thus the multiply-connected regions of Fig. D.8b, c have respectively one and three holes in them.

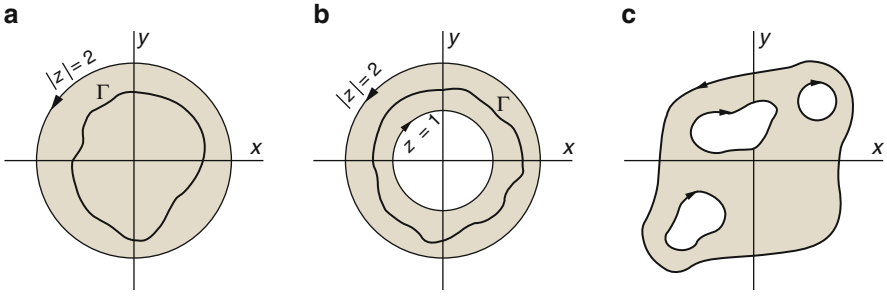


Fig. D.8 Region $\mathfrak{R}$ Curve

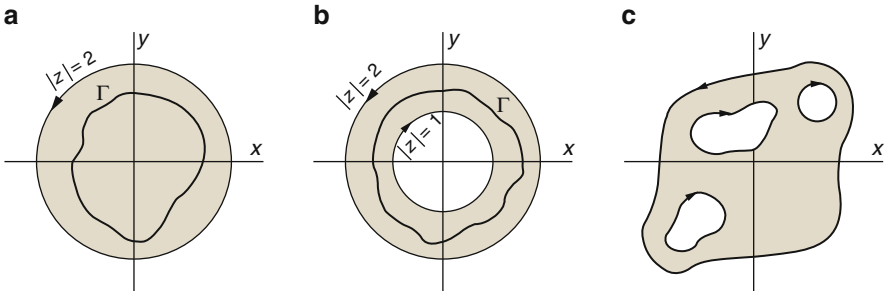


Fig. D.9 Contour Configuration

D.6.7 Convention Regarding Traversal of a Closed Path

The boundary C of a region is said to be traversed in the *positive sense* or *direction* if an observer traveling in this direction (and perpendicular to the plane) has the region to the left. This convention leads to the direction indicated by the arrows in Fig. D.9a–c. We use the special symbol

$$\oint_C f(z)dz$$

To denote integration of $f(z)$ around the boundary C in the positive sense. Note that in the case of circle (Fig. D.9a) the positive direction is the *counterclockwise direction*. The integral around C is often called a *contour integral*.

D.6.8 Cauchy's Theorem

Let $f(z)$ be analytic in region $\mathfrak{R}$ and on its boundary C. Then

$$\oint_C f(z)dz = 0 \tag{D.9}$$

This fundamental theorem, often called *Couchy's integral theorem* or briefly *Couchy's theorem*, is valid for both simply and multiply-connection regions. It was first proved by use of Green's theorem with the added restriction that $f'(z)$ be continuous in $\mathfrak{R}$ (See Example 1 below). However, *Goursat* gave a proof which removed this restriction. For this reason the theorem is sometimes called *Cauchy–Goursat theorem* (See Example 2 to Example 5) when one desires to emphasize the removal of this restriction.

Example 1: Prove Cauchy's theorem $\oint_C f(z)dz = 0$ if $f(z)$ is analytic with derivative $f'(z)$ which is continuous at all points inside and on a simple closed curve C.

Solution: If $f(z) = u + iv$ is analytic and has a continuous derivative

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial v}{\partial y}$$

it follows that the partial derivatives (1) $\dfrac{\partial u}{\partial x} = \dfrac{\partial v}{\partial y}$, (2) $\dfrac{\partial v}{\partial x} = -\dfrac{\partial v}{\partial y}$ are continuous inside and on C. Thus Green's theorem can be applied and we have

$$\oint_C f(z)dz = \oint_C (u + iv)(dx + idy) = \oint_C udx - vdy + i\oint_C vdx + udy$$

$$= \iint_{\mathfrak{R}} \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)dxdy + i\iint_{\mathfrak{R}} \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)dxdy = 0$$

Using Cauchy-Riemann equation (1) and (2) in above.

By using the fact that Green's theorem is applicable to multiply connected regions, we can extend the result to multiply-connected regions under the given conditions on $f(z)$. The Cauchy–Goursat theorem (See Example 2 to 3) removes the restriction that $f'(z)$ be continuous.

Other Methods.

The result can be obtained from the complex form of Green's theorem Example 3 in Sect. 22.9 in above by nothing that if $B(z, \overline{z}) = f(z)$ is independent of $\overline{z}$, then $\partial B/\partial \overline{z} = 0$ and so $\oint_C f(z)dz = 0$.

D.6.9 Morera's Theorem

Let $f(z)$ be continuous in a simply-connected region $\mathfrak{R}$ and suppose that

$$\oint_c f(z)dz = 0 \tag{D.10}$$

around *every* simple closed curve C in $\mathfrak{R}$. Then $f(z)$ is analytic in $\mathfrak{R}$.

This theorem, due to *Morera*, is often called the *converse of Cauchy's theorem*. It can be extended to multiply-connected regions. For a proof which assumes that $f'(z)$ is continuous in $\mathfrak{R}$, see the following Example 1. For a proof which eliminates this restriction, see Example 2 below as well.

Example 1: Prove Morera's theorem under the assumption that $f(z)$ has a continuous derivative in $\mathfrak{R}$.

Solution: If $f(z)$ has a continuous derivative in $\mathfrak{R}$, then we can apply Green's theorem to obtain

$$\oint_c f(z)dz = \oint_c udx - vdy + i\oint_c vdx - udy$$
$$= \int\int_c \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)dxdy + i\int\int_c \left(-\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)dxdy$$

Then if $\oint_c f(z)dz = 0$ around every closed path C in $\mathfrak{R}$, we must have

$$\oint_c udx - vdy = 0 \quad \text{and} \quad \oint_c vdx - udy = 0$$

around every closed path C in $\mathfrak{R}$.

Example 2: If $f(z)$ is analytic in a simple-connected region $\mathfrak{R}$, prove that $\int_a^b f(z)dz$ is independent of the path in $\mathfrak{R}$ joining any two points a and b in $\mathfrak{R}$. (Fig. D.10)

Solution: By Cauchy's theorem,

$$\int_{\text{ADBEA}} f(z)dz = 0$$

or

$$\int_{\text{ADB}} f(z)dz + \int_{\text{BEA}} f(z)dz = 0$$

Hence

$$\int_{\text{ADB}} f(z)dz = -\int_{\text{BEA}} f(z)dz = \int_{\text{AEB}} f(z)dz$$

Thus

$$\int_{C_1} f(z)dz = \int_{C_2} f(z)dz = \int_a^b f(z)dz$$

which yields the required result

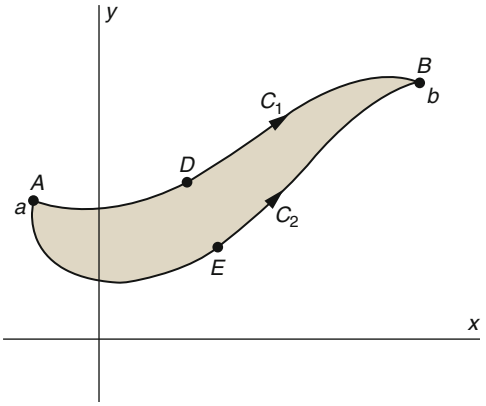


Fig. D.10 Contour Curve C_1 and C_2

Example 3: Prove *Morera's theorem* (the converse of *Cauchy's theorem*): If $f(z)$ is continuous in a simply-connected region $\mathfrak{R}$ and if $\oint_C f(z)dz = 0$ around every simple closed curve C in region $\mathfrak{R}$, then $f(z)$ is analytic in $\mathfrak{R}$.

Solution: If $\oint_C f(z)dz = 0$ independent of C, it follows by Example 2 above, that $F(z) = \int_a^z f(z)dz$ is independent of the path joining a and z, so long as this path is in $\mathfrak{R}$. Then by reasoning identical with that used in Example 3 in above, it follows that $F(z)$ is analytic in $\mathfrak{R}$ and $F'(z) = f(z)$. However, by Example 2 of Sect. 18, it follows that $F'(z)$ is also analytic if $F(z)$ is. Hence $f(z)$ is analytic in $\mathfrak{R}$.

Example 4: Let $f(z)$ be analytic in region $\mathfrak{R}$ bounded by two simple closed curves C_1 and C_2 [See Fig. D.11] and also C_1 and C_2. Prove that $\oint_{C_1} f(z)dz = \oint_{C_2} f(z)dz$, where C_1 and C_2 are both traversed in the positive sense relative to their interiors [counter-clockwise in Fig. D.36].

Solution: Consider cross-cut DE. Then since $f(z)$ is analytic in the region $\mathfrak{R}$, we have by Cauchy's theorem

$$\int_{\text{DEFGEDHJKLD}} f(z)dz = 0$$

$$\text{or} \quad \int_{\text{DE}} f(z)dz + \int_{\text{EFGE}} f(z)dz + \int_{\text{ED}} f(z)dz + \int_{\text{DHJKLD}} f(z)dz = 0$$

$$\text{Hence since} \quad \int_{\text{DE}} f(z)dz = -\int_{\text{ED}} f(z)dz$$

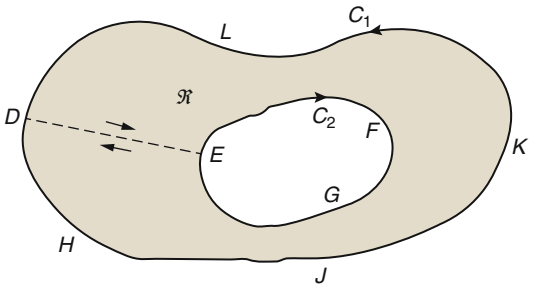


Fig. D.11 Path of Contour

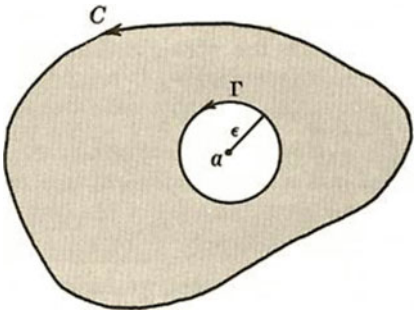


Fig. D.12 Contour Configuration

$$\int_{\text{DHJKLD}} f(z)dz = -\int_{\text{EFGE}} f(z)dz = \int_{\text{EGFE}} F(z)dz \quad \text{or}$$

Example 5: Evaluate $\oint_C \frac{dz}{z-a}$ where C is any simple closed curve C and $z = a$ is

(a) Outside C.
(b) Inside C.

Solution:

(a) If a is outside C, then $f(z) = 1/(z-a)$ is analytic everywhere inside and on C. Hence by Cauchy's theorem $\oint_C \frac{dz}{z-a} = 0$ (Fig. D.12)

(b) Suppose a is inside C and let Γ be a circle of radius ε with center at $z = a$ so that Γ is inside C [this can be done since $z = a$ is an interior point]. By Example 4 above we have

$$\oint_C \frac{dz}{z-a} = \oint_\Gamma \frac{dz}{z-a} \tag{D.11}$$

Now on Γ, $|z-a|=\varepsilon$ or $z-a=\varepsilon e^{i\theta}$, i.e. $z=a+\varepsilon e^{i\theta}$, $0\le\theta<2\pi$. Thus since $dz=i\varepsilon e^{i\theta}d\theta$, the right side of Eq. (D.11) becomes

$$\int_{\theta=0}^{2\pi}\frac{i\varepsilon e^{i\theta}d\theta}{\varepsilon e^{i\theta}}=i\int_{\theta=0}^{2\pi}d\theta=2\pi i$$

which is the required value.

Example 6: Evaluate $\oint_C \frac{dz}{(z-a)^n}$ $n=2,3,4\cdots$ where $z=a$ is inside the simple closed curve C.

Solution: As Example 5 of above

$$\begin{aligned}\oint_C\frac{dz}{(z-a)^n}&=\oint_\Gamma\frac{dz}{(z-a)^n}\\&=\int_{\theta=0}^{2\pi}\frac{i\varepsilon e^{i\theta}d\theta}{\varepsilon^n e^{in\theta}}=\frac{i}{\varepsilon^{n-1}}\int_{\theta=0}^{2\pi}e^{(1-n)i\theta}d\theta\\&=\frac{i}{\varepsilon^{n-1}}\frac{e^{(1-n)i\theta}}{\varepsilon^n e^{in\theta}}\Bigg|_0^{2\pi}=\frac{1}{(1-n)\varepsilon^{n-1}}\left[\varepsilon^{2(n-1)\pi i}-1\right]=0\end{aligned}$$

where $n\neq 1$.

D.6.10 Some Consequences of Cauchy's Theorem

Let $f(z)$ be analytic in a simply-connected region $\mathfrak{R}$. Then the following theorems hold.

Theorem 1. *If a and z are any two points in* $\mathfrak{R}$, *then*

$$\int_a^z f(z)dz \tag{D.12}$$

Is *independent of the path* in $\mathfrak{R}$ joining a and z.

Theorem 2. *If a and z are any two points in* $\mathfrak{R}$ *and*

$$G(z)=\int_a^b f(z)dz \tag{D.13}$$

Then $G(z)$ is analytic in $\mathfrak{R}$ and $G'(z) = f(z)$. Sometime above expression can be defined as the following form as well;

$$G(z) = \int_a^z f(\xi)d\xi \tag{D.14}$$

Theorem 3. *If a and b are any two points in* $\mathfrak{R}$ *and* $F'(z) = f(z)$, *then*

$$\int_a^b f(z)dz = F(b) - F(a) \tag{D.15}$$

This can also be written in the form, familiar from elementary calculus,

$$\int_a^b f(z)dz = F(z)\Big|_a^b = F(b) - F(a) \tag{D.16}$$

$$\text{Example}: \quad \int_{3i}^{1-i} 4zdz = 2z^2\Big|_{3i}^{1-i} = 2(1-i)^2 - 2(3i)^2 = 18 - 4i$$

Theorem 4. *If f(z) be analytic in a region bounded by two simple closed curves C and* C_1 *[where* C_1 *lies inside C as in Fig. D.13 below] and on these curves. Then*

$$\oint_C f(z)dz = \oint_{C_1} f(z)dz \tag{D.17}$$

where C and C_1 are both traversed in the positive sense relative to their interiors [counter-clockwise in Fig. D.14].

The result shows that if we wish to integrate $f(z)$ along curve C we can equivalently replace C by any curve C_1 so long as $f(z)$ is analytic in the region between C and C_1.

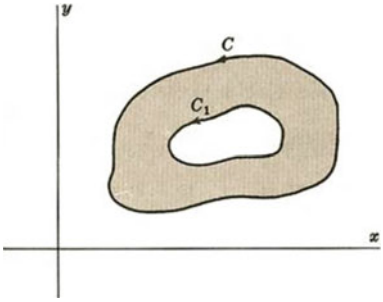


Fig. D.13 Contour Configuration

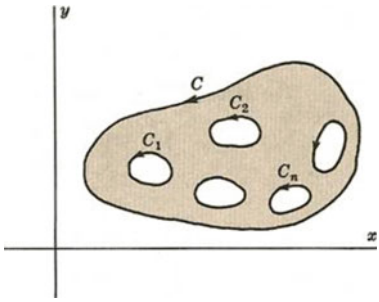


Fig. D.14 Contour Configuration

Theorem 5. *If f(z) be analytic in a region bounded by the non-overlapping simple closed curves* $C, C_1, C_2, C_3, \ldots C_n$ *[where* $C, C_1, C_2, C_3, \ldots C_n$ *are inside C as in Fig. D.12 above] and on these curves. Then*

$$\oint_C f(z)dz = \oint_{C_1} f(z)dz + \oint_{C_2} f(z)dz + \oint_{C_3} f(z)dz + \ldots + \oint_{C_n} f(z)dz \tag{D.18}$$

This is a generalization of the Theorem 4.

D.7 Cauchy's Integral Formulas and Related Theorem

In this section we will talk about Cauchy's integral formula that states where the integral is a contour integral along the contour γ enclosing the point a and related other theorem

D.7.1 Cauchy's Integral Formulas

If $f(z)$ is analytic inside and on a simple closed curve C and z_0 is any point inside C (Fig. D.15), then

$$f(z_0) = \frac{1}{3\pi i}\oint_\gamma \frac{f(z)}{z - z_0}dz \tag{D.19}$$

where the integral is a contour integral along the contour γ enclosing the point z_0. It can be derived by considering the contour integral

$$\oint_\gamma \frac{f(z)}{z - z_0}dz \tag{D.20}$$

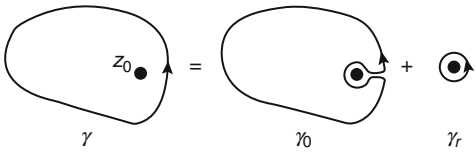


Fig. D.15 Contour Configuration

defining a path γ_r as an infinitesimal counterclockwise circle around the point z_0, and defining the path γ_0 as an arbitrary loop with a cut line (on which the forward and reverse contributions cancel each other out) so as to go around z_0. The total path is then

$$\gamma = \gamma_0 + \gamma_r \tag{D.21}$$

so

$$\oint_{\gamma} \frac{f(z)}{z - z_0} dz = \oint_{\gamma_0} \frac{f(z)}{z - z_0} dz + \oint_{\gamma_r} \frac{f(z)}{z - z_0} dz \tag{D.22}$$

From the Cauchy integral theorem, the contour integral along any path not enclosing a pole is 0. Therefore, the first term in the above equation is 0 since γ_0 does not enclose the pole, and we are left with

$$\oint_{\gamma} \frac{f(z)}{z - z_0} dz = \oint_{\gamma_r} \frac{f(z)}{z - z_0} dz \tag{D.23}$$

Now, let $z = z_0 + re^{i\theta}$, so $dz = ire^{i\theta} d\theta$. Then

$$\begin{aligned} \oint_{\gamma} \frac{f(z)}{z - z_0} dz &= \oint_{\gamma_r} \frac{f\left(z_0 + re^{i\theta}\right)}{z - z_0} ire^{i\theta} d\theta \\ &= \oint_{\gamma_r} \left(z_0 + re^{i\theta}\right) id\theta \end{aligned} \tag{D.24}$$

But we are free to allow the radius r to shrink to 0, so

$$\begin{aligned} \oint_{\gamma} \frac{f(z)}{z - z_0} dz &= \lim_{x \to 0} \oint_{\gamma_r} \frac{f\left(z_0 + re^{i\theta}\right)}{z - z_0} id\theta \\ &= \oint_{\gamma_r} f(z_0) id\theta \\ &= if(z_0) \oint_{\gamma_r} d\theta \\ &= 2\pi i f(z_0) \end{aligned} \tag{D.25}$$

Given Eq. D.25.

If multiple loops are made around the point z_0, then Eq. D.25 becomes

$$f(z_0) = \frac{1}{2\pi i}\oint_\gamma \frac{f(z)}{z - z_0} dz$$

A similar formula holds for the derivative of $f(z)$

$$\begin{aligned} f'(z_0) &= \lim_{h\to 0} \frac{f(z_0 + h) - f(z_0)}{h} \\ &= \lim_{h\to 0} \frac{1}{2\pi i h}\left[\oint_\gamma \frac{f(z)dz}{z - z_0 - h} - \oint_\gamma \frac{f(z)dz}{z - z_0}\right] \\ &= \lim_{h\to 0} \frac{1}{2\pi i h}\left[\oint_\gamma \frac{f(z)[(z - z_0) - (z - z_0 - h)]dz}{(z - z_0 - h)(z - z_0)}\right] \\ &= \lim_{h\to 0} \frac{1}{2\pi i h}\oint_\gamma \frac{hf(z)dz}{(z - z_0 - h)(z - z_0)} \\ &= \lim_{h\to 0} \frac{1}{2\pi i}\oint_\gamma \frac{f(z)dz}{(z - z_0)^2} \end{aligned} \tag{D.26}$$

Iteration again,

$$f''(z_0) = \frac{2}{2\pi i}\oint_\gamma \frac{f(z)dz}{(z - z_0)^3} \tag{D.27}$$

Continuing the process n times, we have

$$f^n(z_0) = \frac{n!}{2\pi i}\oint_\gamma \frac{f(z)dz}{(z - z_0)^{n+1}} \quad n = 1, 2, 3, \ldots$$

Example 1: If $f(z)$ is analytic inside and on boundary C of a simply-connected region $\mathfrak{R}$, prove Cauchy's integral formula

$$f(a) = \frac{1}{2\pi i}\oint_C \frac{f(z)}{z - a} dz$$

Solution: Method 1. The function $f(z)/(z - a)$ is analytic inside and on C except at the point $z = a$ (See Fig. D.16). By theorem 4 of Sect. 22.14, we have

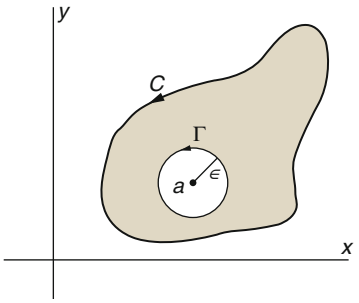


Fig. D.16 Contour Configuration

$$\oint_C \frac{dz}{z-a} = \oint_\Gamma \frac{dz}{z-a} \tag{D.28}$$

where we can chosen Γ as a circle of radius ε with center at a. Then an equation for Γ is $|z-a| = \varepsilon$ or $z - a = \varepsilon e^{i\theta}$ where $0 \le \theta < 2\pi$. Substituting $z = a + \varepsilon e^{i\theta}$, $dz = i\varepsilon e^{i\theta} d\theta$, the integral on the right of Eq. (D.28) becomes

$$\oint_\Gamma \frac{f(z)}{z-a} dz = \int_{\theta=0}^{2\pi} \frac{f\left(a+\varepsilon e^{i\theta}\right) i\varepsilon e^{i\theta}}{\varepsilon e^{i\theta}} d\theta = i\int_{\theta=0}^{2\pi} f\left(a+\varepsilon e^{i\theta}\right) d\theta$$

Thus we have from Eq. (D.28)

$$\oint_C \frac{f(z)}{z-a} dz = i\int_{\theta=0}^{2\pi} f\left(a+\varepsilon e^{i\theta}\right) d\theta \tag{D.29}$$

Taking the limits of both sides of Eq. (D.29) and making use of the continuity of $f(z)$, we have

$$\begin{aligned} \oint_C \frac{f(z)}{z-a} dz &= \lim_{\varepsilon\to 0} i\int_{\theta=0}^{2\pi} f\left(a+\varepsilon e^{i\theta}\right) d\theta \\ &= i\int_{\theta=0}^{2\pi} \lim_{\varepsilon\to 0} f\left(a+\varepsilon e^{i\theta}\right) d\theta = i\int_{\theta=0}^{2\pi} f(a) d\theta = 2\pi i f(a) \end{aligned} \tag{D.30}$$

So that we have, as required,

$$f(a) = \frac{1}{2\pi i}\int_C \frac{f(z)}{z-a} dz$$

Method 2. The right side of Eq. (D.28) of Method 1 can be written as

$$\oint_\Gamma \frac{f(z)}{z-a}dz = \oint_\Gamma \frac{f(z)-f(a)}{z-a}dz + \oint_\Gamma \frac{f(z)}{z-a}dz$$

$$= \oint_\Gamma \frac{f(z)-f(a)}{z-a}dz + 2\pi i f(a)$$

using Example 5 of Sect. 22.11 in above. The required result will follow if we can show that

$$\oint_\Gamma \frac{f(z)-f(a)}{z-a}dz = 0$$

But by Example 1 of Sect. 3.4 we have

$$\oint_\Gamma \frac{f(z)-f(a)}{z-a}dz = \oint_\Gamma f'(a)dz + \oint_\Gamma \eta dz = \oint_\Gamma \eta dz$$

Then choosing Γ so small that for all points on Γ we have $|\eta| < \delta/2\pi$, we find

$$\left|\oint_\Gamma \eta dz\right| < \left(\frac{\delta}{2\pi}\right)(2\pi\varepsilon) = \varepsilon$$

Thus $\oint_\Gamma \eta dz = 0$ and the proof is complete.

Example 2: If $f(z)$ is analytic inside and on the boundary C of a simply-connected region $\mathfrak{R}$, prove that

$$f'(a) = \frac{1}{2\pi i}\oint_\Gamma \frac{f(z)}{(z-a)^2}dz$$

Solution: From Example 1 in above (Sect. 23.1) if a and $a+h$ lie in $\mathfrak{R}$, we have

$$\frac{f(a+h)-f(a)}{h} = \frac{1}{2\pi i}\oint_C \frac{1}{h}\left\{\frac{1}{z-(a+h)} - \frac{1}{z-a}\right\}f(z)dz$$

$$= \frac{1}{2\pi i}\oint_C \frac{f(z)dz}{(z-a-h)(z-a)} = \frac{1}{2\pi i}\oint_C \frac{f(z)dz}{(z-a)^2}$$

$$+\frac{h}{2\pi i}\oint_C \frac{f(z)dz}{(z-a-h)(z-a)^2}$$

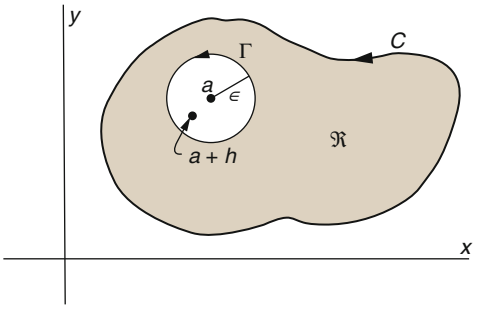


Fig. D.17 Contour Configuration

The result follows on taking the limit as $h \to 0$ if we can show that the last term approaches zero.

To show this we use the fact that if Γ is a circle of radius ε and center a which lies in $\mathfrak{R}$ (See Fig. D.17), then

$$\frac{h}{2\pi i}\oint_C \frac{f(z)dz}{(z-a-h)(z-a)^2}$$

$$\frac{h}{2\pi i}\oint_\Gamma \frac{f(z)dz}{(z-a-h)(z-a)^2}$$

Choosing h so small in absolute value that $a+h$ lies in Γ and $|\boldsymbol{h}| < \varepsilon/2$, we have by argument in Sect. 1.4 ($|z_1 - z_2| \geq |z_1| - |z_2|$), and the fact that Γ has equation $|z-a| = \varepsilon$,

$$|z-a-h| \geq |z-a| - |h| > \varepsilon - \varepsilon/2 = \varepsilon/2$$

Also since $f(z)$ is analytic in $\mathfrak{R}$ we can find a positive number M such that $|f(z)| < M$.

Then since the length of Γ is $2\pi\varepsilon$, we have

$$\left|\frac{h}{2\pi i}\oint_\Gamma \frac{f(z)dz}{(z-a-h)(z-a)^2}\right| \leq \frac{|h|}{2\pi}\frac{M(2\pi\varepsilon)}{(\varepsilon/2)(\varepsilon^2)} = \frac{2|h|M}{\varepsilon^2}$$

and if follows that the left side approaches zero as $h \to 0$, thus completing the proof. It is of interest to observe $h \to 0$ that the result is equivalent to

$$\frac{d}{dt}f(a) = \frac{d}{da}\left\{\frac{1}{2\pi i}\oint_C \frac{f(z)}{z-a}dz\right\} = \frac{1}{2\pi i}\oint_C \frac{\partial}{\partial z}\left\{\frac{f(z)}{z-a}\right\}dz$$

which is an extension to contour integrals of *Laibmitz's rul*e for differentiating under the integral sign.

Example 3: Prove that under the condition of Example 2

$$f^{(n)}(a) = \frac{n!}{2\pi i}\oint_C \frac{f(z)}{(z-a)^{n+1}}dz \quad n = 0, 1, 2, 3, \ldots$$

The case where $n = 0$ and 1 follow from Example 1 and 2 respectively provided we define $f^{(0)}(a) = f(a)$ and $0! = 1$.

Solution: To establish the case where $n = 2$, we use problem 2 where a and $a + h$ lie in $\mathfrak{R}$ to obtain

$$\begin{aligned}\frac{f'(a+h) - f'(a)}{h} &= \frac{1}{2\pi i}\oint_C \left\{\frac{1}{(z-a-h)^2} - \frac{1}{(z-a)^2}\right\} f(z)dz \\ &= \frac{2!}{2\pi i}\oint_C \frac{1}{(z-a)^3} f(z)dz + \frac{h}{2\pi i}\oint_C \frac{3(z-a)-2h}{(z-a-h)^2(z-a)^3} f(z)dz\end{aligned}$$

The result follows on taking the limit as $h \to 0$ if we can show that the last term approaches zero. The proof is similar to that of Example 2, for using the fact that the integral around C equals the integral around Γ, we have

$$\left|\frac{2}{2\pi i}\oint_\Gamma \frac{3(z-a)-2h}{(z-a-h)^2(z-a)^3} f(z)dz\right| \le \frac{|h|}{2\pi}\frac{M(2\pi)}{(\varepsilon/2)^2(\varepsilon^3)} = \frac{4|h|M}{\varepsilon^4}$$

Since M exists such that $\left|\{3(z-a)-2h\}f(z)\}\right| < M$.

In a similar manner we can establish the result for $n = 3, 4, \ldots$.

The Result is equivalent to last paragraph of Example-2

$$\frac{d^n}{da^n} f(a) = \frac{d^n}{da^n}\left\{\frac{1}{2\pi i}\oint_C \frac{f(z)}{(z-a)}dz\right\} = \frac{1}{2\pi i}\oint_C \frac{\partial^n}{a^n}\left\{\frac{f(z)}{(z-a)}\right\}dz$$

Example 4: if $f(z)$ is analytic in region $\mathfrak{R}$, prove that $f'(z), f''(z), \ldots$ are analytic in $\mathfrak{R}$.

Solution: This follows from Example-2 and 3

Example 5: Evaluate

(a) $\oint_C \dfrac{\sin \pi z^2 + \cos \pi z^2}{(z-1)(z-2)}dz$

(b) $\oint_C \dfrac{e^{2z}}{(z+1)^4}dz$ where C is the circle $|z| = 3$

Solution (a) Since $\frac{1}{(z-1)(z-2)} = \frac{1}{(z-2)} - \frac{1}{(z-1)}$, we have

$$\oint_C \frac{\sin \pi z^2 + \cos \pi z^2}{(z-1)(z-2)} dz = \oint_C \frac{\sin \pi z^2 + \cos \pi z^2}{(z-2)} dz - \oint_C \frac{\sin \pi z^2 + \cos \pi z^2}{(z-1)} dz$$

By Cauchy's integral formula with $a = 2$ and $a = 1$ respectively, we have

$$\oint_C \frac{\sin \pi z^2 + \cos \pi z^2}{(z-2)} dz = 2\pi i \left\{ \sin \pi (2)^2 + \cos \pi (2)^2 \right\} = 2\pi i$$

$$\oint_C \frac{\sin \pi z^2 + \cos \pi z^2}{(z-1)} dz = 2\pi i \left\{ \sin \pi (1)^2 + \cos \pi (1)^2 \right\} = -2\pi i$$

Since $z = 1$ and $z = 2$ are inside C and $\sin \pi z^2 + \cos \pi z^2$ is analytic inside C. Then the required integral has the value $2\pi i - (-2\pi i) = 4\pi i$.

(b) Let $f(z) = e^{2z}$ and $a = -1$ in the Cauchy integral formula

$$f^{(n)}(a) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z-a)^{n+1}} dz \tag{D.31}$$

If $n = 3$, the $f'''(z) = 8e^{2z}$ and $f'''(-1) = 8e^{-2}$. Hence Eq. (D.31) becomes

$$8e^{-2} = \frac{3!}{2\pi i} \oint_C \frac{e^{2z}}{(z+1)^4} dz$$

From which we see that the required integral has the value $8\pi i e^{-2}/3$.

Example 6: Prove Cauchy's integral formula for multiply-connected regions

Solution: We present a proof for the multiply-connected region $\mathfrak{R}$ bounded by the simple closed curves C_1 and C_2 as indicated in Fig. D.18. Extensions to other multiply-connected regions are easily made.

Construct a circle Γ having center at any point a in $\mathfrak{R}$ so that Γ lies entirely in $\mathfrak{R}$. Let $\mathfrak{R}'$ consist of the set of points in $\mathfrak{R}$ which are exterior to Γ. Then the function $\frac{f(z)}{z-a}$ is analytic inside and on the boundary of $\mathfrak{R}'$.

Hence by Cauchy's theorem for multiply-connected regions

$$\frac{1}{2\pi i} \oint_{C_1} \frac{f(z)}{z-a} dz - \frac{1}{2\pi i} \oint_{C_2} \frac{f(z)}{z-a} dz - \frac{1}{2\pi i} \oint_{\Gamma} \frac{f(z)}{z-a} dz = 0 \tag{D.32}$$

But by Cauchy's integral formula for simply-connected regions, we have

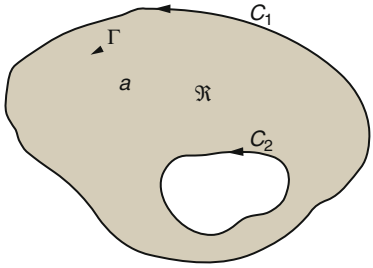


Fig. D.18 Contour Configuration

$$f(a) = \frac{1}{2\pi i}\oint_{\Gamma} \frac{f(z)}{z-a}dz \tag{D.33}$$

So that from (D.32) we have

$$f(a) = \frac{1}{2\pi i}\oint_{C_1} \frac{f(z)}{z-a}\ dz - \frac{1}{2\pi i}\oint_{C_1} \frac{f(z)}{z-a}\ dz \tag{D.34}$$

Then if C represents the entire boundary of $\Re$ (suitably traversed so that an observer moving around C always has $\Re$ lying to his left), we can write Eq. (D.34) as

$$f(a) = \frac{1}{2\pi i}\oint_{C} \frac{f(z)}{z-a}dz$$

In a similar manner we can show that the other Cauchy integral formulas

$$f^{(n)}(a) = \frac{n!}{2\pi i}\oint_{C} \frac{f(z)}{(z-a)^{n+1}}dz \quad n = 0, 1, 2, 3, \ldots$$

holds for multiply-connected regions.

D.8 Classification of Singularities

If is possible to classify the singularities of a function $f(z)$ by examination of its Laurent series. For this purpose we assume that in Fig. D.18, $R_2 = 0$, so that $f(z)$ is analytic inside and on C_1 except at $z = a$ which is an isolated unless otherwise indicated.

D.8.1 Poles

If $f(z)$ has the form Eq. (D.5) Sect. 26.1 in above in which the principal part has only finite number of terms given by

$$\frac{a_{-1}}{z-a}+\frac{a_{-2}}{(z-a)^2}+\cdots+\frac{a_{-n}}{(z-a)^n}$$

where $a_{-n}\neq 0$, then $z=a$ is called a *pole of order n*. If $n=1$, it is called a *simple pole*.

If $f(z)$ has a pole at $z=a$, then $\lim_{x\to a} f(z)=\infty$. See Example 1 below.

Example 1: Prove that an analytic function cannot be bounded in the neighborhood of an isolated singularity.

Solution: Let $f(z)$ be analytic inside and on a circle C of radius r, except at the isolated singularity $z=a$ taken to be the center of C. Then by Laurent's Theorem $f(z)$ has a Laurent Expansion of,

$$f(z)=\sum_{k=-\infty}^{\infty} a_k(z-a)^k \tag{D.35}$$

where the coefficients a_k are given by Eq. (D.34) of Sect. 26.1 in above. In particular,

$$a_{-n}=\frac{1}{2\pi i}\oint_C \frac{f(z)}{(z-a)^{-n+1}}dz \quad n=1,2,3,\ldots \tag{D.36}$$

Now if $|f(z)|<M$ for a constant M, i.e., if $f(z)$ is bounded, then from Eq. (D.36) above,

$$a_{-n}=\frac{1}{2\pi i}\left|\oint_C \frac{f(z)}{(z-a)^{-n+1}}dz\right|\leq \frac{1}{2\pi}r^{n-1}\cdot M\cdot 2\pi r=Mr^n$$

Hence since r can be made arbitrarily small, we have $a_{-n}=0$, $n=1,2,3,\ldots$, i.e. $a_{-1}=a_{-2}=a_{-3}=\cdots=0$, and the Laurent series reduces to a Taylor series about $z=a$. This shows that $f(z)$ is analytic at $z=a$ so that $z=a$ is not a singularity, contrary to hypothesis. This contradiction shows that $f(z)$ cannot be bounded in the neighborhood of an isolated singularity.

D.8.2 Removable Singularities

If a single-valued function $f(z)$ is not defined at $z = a$ but $\lim_{x\to a} f(z)$ exists, then $z = a$ is called a *removable singularity*. In such case we define $f(z)$ at $z = a$ as equal to $\lim_{x\to a} f(z)$.

Example 1: if $f(z) = \dfrac{\sin z}{z}$, then $z = 0$ is a removable singularity $f(0)$ is not defined but $\lim_{x\to 0} \dfrac{\sin z}{z} = 1$. We define $f(0) = \lim_{x\to 0} \dfrac{\sin z}{z} = 1$. Note that in this case

$$\frac{\sin z}{z} = \frac{1}{z}\left\{z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots\right\} = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \frac{z^6}{7!} + \cdots$$

D.8.3 Essential Singularities

If $f(z)$ is single-valued, then any singularity which is not a pole or removable singularity is called an *essential singularity*. If $z = a$ is an essential singularity of $f(z)$, the principal part of the Laurent expansion has infinitely many terms.

Example 1: Since $e^{1/z} = 1 + \frac{1}{z} + \frac{1}{2!z^2} + \frac{1}{3!z^3} + \cdots$, $z = 0$ is an essential singularity.

D.8.4 Brach Point

A point $z = z_0$ is called a *branch point* of the multiple-valued function $f(z)$ if the branches of $f(z)$ are interchanged when z describes a closed path about z_0. Since each of the branches of a multiple-valued function is analytic, all of the theorem for analytic functions, in particular Taylor's Theorem.

Example 1: The branch of $f(z) = z^{1/2}$ which has the value 1 for $z = 1$, has a Taylor series of the form $a + a_1(z-1) + a_2(z-a)^2 + \cdots$ with radius of convergence $R = 1$ [the distance from $z = 1$ to the nearest singularity namely the branch point $z = 0$].

D.8.5 Singularities at Infinity

By letting $z = 1/w$ in $f(z)$, we obtain the function $f(1/w) = F(w)$. Then the nature of singularity at $z = \infty$ [the point at infinity] is defined to be the same as that of $F(w)$ at $w = 0$.

Example 1: $f(z) = z^2$ has a pole of order 3 at $z = \infty$, since $F(w) = f(1/w) = 1/w^2$ has a pole of order 3 at $w = 0$. Similarly $f(z) = e^z$ has an essential singularity at $z = \infty$, since $F(w) = f(1/w) = e^{1/w}$ has an essential singularity at $w = 0$.

D.9 Entire Functions

A function which is analytic everywhere in the finite plane [i.e., everywhere except at ∞] is called an *entire function* or *integral function*. The functions e^z, $\sin z$ and $\cos z$ are entire functions.

An entire function can be represented by a Taylor Series (See Appendix A) which has an infinite radius of convergence. Conversely if a power series has an infinite radius of convergence, it represents an entire function.

Note that by Liouville's theorem a function which is analytic *everywhere including* ∞ must be a constant.

D.10 The Residues Theorem, Evaluation of Integrals

In the mathematical field of complex analysis, contour integration is a method of evaluating certain integrals along paths in the complex plane.

Contour integration is closely related to the calculus of residues,[4] a methodology of complex analysis.

One use for contour integrals is the evaluation of integrals along the real line that are not readily found by using only real variable methods.[5]

Contour integration methods include

- Direct integration of a complex-valued function along a curve in the complex plane (a contour)
- Application of the Cauchy integral formula
- Application of the residue theorem.

One method can be used, or a combination of these methods or various limiting processes, for the purpose of finding these integrals or sums.

D.10.1 Residues

Let $f(z)$ be single-valued and analytic inside and on a circle C except at the point $z = a$ chosen as the center of C. Then, as we have seen in pervious sections (Laurent Theorem and Series) $f(z)$ has a Laurent series about $z = a$ given by

$$\begin{aligned} f(z) &= \sum_{n=-\infty}^{\infty} a_n (z-a)^n \\ &= a_0 + a_1(z-a) + a_2(z-a)^2 + \cdots + \frac{a_{-1}}{z-a} + \frac{a_{-2}}{(z-a)^2} + \cdots \end{aligned} \tag{D.37}$$

$$\text{where} \quad a_n = \frac{1}{2\pi i} \oint_C \frac{f(z)}{(z-a)^{n+1}} dz \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{D.38}$$

In the special case $n = -1$, we have from Eq. D.38

$$\oint_C f(z) dz = 2\pi i a_{-1} \tag{D.39}$$

Formally we obtained Eq. D.39 from 37 by integrating term by term and using the results (Example 5 and Example 6 Sect. D.6.9)

$$\oint_C \frac{f(z)}{(z-a)^p} dz = \begin{cases} 2\pi i & p = 1 \\ 0 & p = \text{integer} \neq 1 \end{cases} \tag{D.40}$$

Because of the fact that Eq. D.39 involves only the coefficient a_{-1} in Eq. D.37, we call a_{-1} the *residue* of $f(z)$ at $z = a$.

D.10.2 Calculation of Residues

To obtain residue of a function $f(z)$ at $z = a$, it may result from Eq. D.37 that the Laurent Expansion of $f(z)$ about $z = a$ must be obtained. However, in the case where $z = a$ is a pole of order k there is a simple formula for a_{-1} given by

$$a_{-1} = \lim_{z \to a} \frac{1}{(k-1)!} \frac{d^{k-1}}{dz^{k-1}} \left\{ (z-a)^k f(z) \right\} \tag{D.41}$$

If $k = 1$ (simple pole) the result is especially simple and is given by

$$a_{-1} = \lim_{z \to a} (z - a) f(z) \tag{D.42}$$

Which is special case of Eq. D.41 with $k = 1$ if we define $0! = 1$.

Example 1: If $f(z) = \frac{z}{(z-1)(z+1)^2}$ then $z = 1$ and $z = -1$ are poles of orders one and two respectively.

Solution: We have, using Eqs. D.41 and D.42 with $k = 2$

Residue at $z = 1$ is $\lim_{z \to 1} (z - 1) \left\{ \frac{z}{(z-1)(z+1)^2} \right\} = \frac{1}{4}$

Residue at $z = -1$ is $\lim_{z \to -1} \frac{1}{1!} \frac{d}{dz} \left\{ (z + 1)^2 \left(\frac{z}{(z-1)(z+1)^2} \right) \right\} = -\frac{1}{4}$

If $z = a$ is an essential singularity, the residue can sometimes be found by using known series expansions.

Example 2: If $f(z) = \frac{z}{(z+2)(z-3)^2}$ then $f(z)$ has two poles: $z = -2$ a pole of order 1, and $z = 3$, a pole of order 2.

Solution: We have, using Eqs. (D.124) and (D.125) with $k = 2$

$$\text{Residue at } z = -2 \text{ is } \lim_{z \to -2} (z + 2) \left\{ \frac{z}{(z + 2)(z - 3)^2} \right\} = -\frac{2}{25}$$

$$\lim_{z \to 3} \frac{1}{1!} \frac{d}{dz} \left\{ (z - 3)^2 \left(\frac{z}{(z + 2)(z - 3)^2} \right) \right\} = -\frac{1}{4}$$

$$\text{Residue at } z = 3 \text{ is } \lim_{z \to 3} \left\{ \frac{1}{z + 2} - \frac{z}{(z + 2)^2} \right\} = -\frac{2}{25}$$

Often the order of the pole will not be known in advance.

Example 3: If $f(z) = e^{-1/z}$ then $z = 0$ is an essential singularity.

Solution: From the known expansion for e^u with $u = -1/z$ we find

$$e^{-1/z} = 1 - \frac{1}{z} + \frac{1}{2!z^2} - \frac{1}{3!z^3} + \cdots$$

from which we see the residue at $z = 0$ is the coefficient of $1/z$ and equals -1.

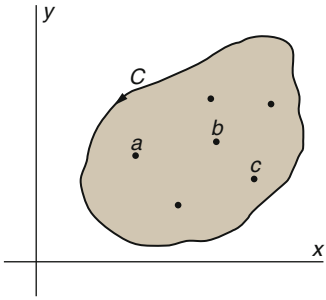


Fig. D.19 Contour Configuration

D.10.3 The Residue Theorem

Let $f(z)$ be single-valued and analytic inside and on a simple closed curve C except at the singularities $a, b, c, \ldots$ inside C which have residues given by a_{-1}, $b_{-1}, c_{-1}, \ldots$ (See Fig. D.19) then the *residues theorem* states that

$$\oint_C f(z) = 2\pi i(a_{-1}, b_{-1}, c_{-1}, +, \ldots) \tag{D.43}$$

i.e., the integral of $f(z)$ around C is $2\pi i$ times the sum of the residues of $f(z)$ at the singularities enclosed by C. Note that Eq. D.43 is a generalization of Eq. D.39. Cauchy's theorem and integral formulas are special cases of this theorem.

D.10.4 Evaluation of Definite Integrals

The evaluation of definite integrals is often achieved by using the residue theorem together with a suitable function $f(z)$ and a suitable closed path or contour C, the choice of which may require great ingenuity. The following types are most common in practice.

Type I: $\int_{-\infty}^{\infty} F(x)dx$, $F(x)$ is a rational function. In this case consider $\oint_C F(z)dz$ along a contour C consisting of the line along the x axis from $-R$ to $+R$ and the semicircle Γ above the x axis having this line as diameter (Fig. D.20). Then let $R \to \infty$. If $F(x)$ is an even function this can be used to evaluate $\int_{-\infty}^{\infty} F(x)dx$. See below Examples

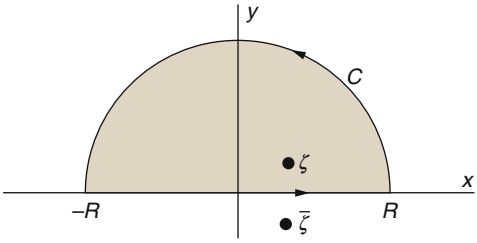


Fig. D.20 Contour Configuration

Example 1: If $\left|\int_C f(z)dz\right| \le ML$ for $z = Re^{i\theta}$ where $k > 1$ and M are constants, prove that $\lim_{R\to\infty}\int_\Gamma F(z)dz = 0$ where Γ is the semicircular arc of radius R shown in Fig. D.20 below.

Solution: If $f(z)$ is integrable along C, then $\left|\int_C f(z)dz\right| \le M$ where $|F(z)| \le M$, i.e. M is an *upper bound* of $|F(z)|$ on C, and L is the *length* of C. With this in mind we have

$\left|\int_\Gamma f(z)dz\right| \le \frac{M}{R^k} \cdot \pi R = \frac{\pi M}{R^{k-1}}$ since the length of arc Γ is $L = \pi R$.

Then $\lim_{R\to\infty}\int_\Gamma F(z)dz = \lim_{R\to\infty}\frac{\pi M}{R^{k-1}} = 0$ and so $\lim_{R\to\infty}\int_\Gamma F(z)dz = 0$

Example 2: Show that for $z = Re^{i\theta}$, $|F(z)| \le M/R^k$, and $k > 1$ if $f(z) = \frac{1}{z^6+1}$.

Solution: If $z = Re^{i\theta}$, then $|f(z)| = \left|\frac{1}{R^6 e^{6i\theta}+1}\right| \le \frac{1}{|R^6 e^{6i\theta}|-1} = \frac{1}{R^6-1} \le \frac{2}{R^6}$ if R is large enough (say $R > 2$, for example) so that $M = 2$ and $k = 6$.

Example 3: Evaluate $\int_0^\infty \frac{dx}{x^6-1}$.

Solution: Consider $\oint_C \frac{dx}{x^6-1}$ where C is the closed contour of Fig. D.21 consisting of the line from $-R$ to R and the semicircle Γ, traversed in the positive (counterclockwise) sense. Since $x^6 - 1 = 0$ when $z = e^{i\pi/6}$, $z = e^{i3\pi/6}$, $z = e^{i5\pi/6}$, $z = e^{i7\pi/6}$, $z = e^{i9\pi/6}$, $z = e^{i11\pi/6}$, these are simple poles of $1/(x^6-1)$. Only the poles $e^{i\pi/6}$, $e^{i3\pi/6}$ and $e^{i5\pi/6}$ lie within C. Then using L'Hospital's rule,

Residue at $e^{i\pi/6} = \lim_{z\to e^{i\pi/6}}\left\{(z - e^{i\pi/6})\frac{1}{z^6+1}\right\} = \lim_{z\to e^{i\pi/6}}\frac{1}{6z^5} = \frac{1}{6}e^{-5i\pi/6}$

Residue at $e^{i3\pi/6} = \lim_{z\to e^{i3\pi/6}}\left\{(z - e^{i3\pi/6})\frac{1}{z^6+1}\right\} = \lim_{z\to e^{i3\pi/6}}\frac{1}{6z^5} = \frac{1}{6}e^{-5i\pi/2}$

Residue at $e^{i5\pi/6} = \lim_{z\to e^{i5\pi/6}}\left\{(z - e^{i5\pi/6})\frac{1}{z^6+1}\right\} = \lim_{z\to e^{i5\pi/6}}\frac{1}{6z^5} = \frac{1}{6}e^{-25i\pi/6}$

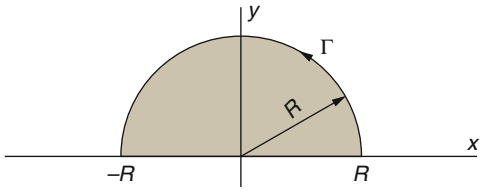


Fig. D.21 Contour Configuration

$$\text{Thus} \oint_C \frac{dx}{x^6-1} = 2\pi i\left\{\frac{1}{6}e^{-5i\pi/6} + \frac{1}{6}e^{-5i\pi/2} + \frac{1}{6}e^{-25i\pi/6}\right\} = \frac{2\pi}{3} \tag{D.44}$$

$$\text{i.e.} \int_{-R}^{R} \frac{dx}{x^6-1} + \int_{\Gamma} \frac{dx}{x^6-1} = \frac{2\pi}{3} \tag{D.45}$$

Taking the limit of both side of Eq. (D.44) as $R \to \infty$ and using Example 1 and Example 2 in above then we have

$$\lim_{R\to\infty} \int_{-R}^{R} \frac{dx}{x^6-1} = \int_{-\infty}^{\infty} \frac{dx}{x^6-1} = \frac{2\pi}{3}$$

Since $\int_{-\infty}^{\infty} \frac{dx}{x^6-1} = 2\int_{0}^{\infty} \frac{dx}{x^6-1}$, then required integral has the value $\pi/3$.

For more examples and solved problem reader should refer to Reference 1 and 2 of this section in particular the suggested one that is written by Murray R. Spiegel where most of these materials are borrowed from.

Type II: $\int_0^{2\pi} G(\sin\theta, \cos\theta)d\theta$**,** $G(\sin\theta, \cos\theta)$ is a rational function of $\sin\theta$ and $\cos\theta$.

Let $z = e^{i\theta}$. The $\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{z - z^{-1}}{2i}$, $\cos\theta = \frac{e^{i\theta} - e^{-i\theta}}{2} = \frac{z + z^{-1}}{2}$ and $dz = ie^{i\theta}d\theta$ or $d\theta = dz/iz$. This given integral is equivalent to $\oint_C F(z)dz$ where C is the unit circle with center at the origin (See Fig. D.22) below. See also the following Examples

Example 1 Evaluate

$$\int_0^{2\pi} \frac{d\theta}{3 - 2\cos\theta + \sin\theta}$$

Solution: Let $z = e^{i\theta}$. Then $\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{z - z^{-1}}{2i}$, $\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2} = \frac{z+z^{-1}}{2}$ and $dz = izd\theta$ so that

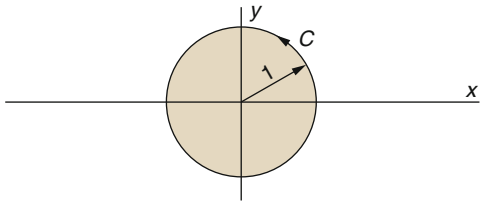


Fig. D.22 Contour Configuration

$$\int_0^{2\pi} \frac{d\theta}{3 - 2\cos\theta + \sin\theta} = \oint_C \frac{dz/iz}{3 - 2(z + z^{-1})/2 + (z - z^{-1})/2i}$$
$$= \oint_C \frac{2dz}{(1 - 2i)z^2 + 6iz - 1 - 2i}$$

where C is the circle of unit radius with center at the origin (See Fig. D.22 above).

The poles of $\frac{2}{(1-2i)z^2+6iz-1-2i}$ are the simple poles

$$z = \frac{-6i \pm \sqrt{(6i)^2 - 4(1 - 20i)(-1 - 2i)}}{2(1 - 2i)}$$
$$= \frac{-6i \pm 4i}{2(1 - 2i)} = 2 - i, \quad (2 - i)/5$$

only $(2 - i)/5$ lies inside C.

by L'Hospital Rule

Residue at

$$(2 - i)/5 = \begin{cases} \lim\limits_{z\to(2-i)/5} \{z - (2 - i)/5\}\left\{\dfrac{2}{(1 - 2i)z^2 + 6iz - 1 - 2i}\right\} \\ \lim\limits_{z\to(2-i)/5} \dfrac{2}{2(1 - 2i)z + 6i} = \dfrac{1}{2i} \end{cases}$$

Then $\oint_C \frac{2dz}{(1 - 2i)z^2 + 6iz - 1 - 2i} = 2\pi i\left(\frac{1}{2i}\right) = \pi$, the required value.

For more examples and solved problem readers should refer to Reference 1 and 2 of this section in particular the suggested one that is written by Murray R. Spiegel where most of these materials borrowed from.

Type III: The integral type $I_1 = \int_{-\infty}^{\infty} R(x)\cos mx dx$ or $I_2 = \int_{-\infty}^{\infty} R(x)\sin mx dx$ where the function $R(x) = P(x)/Q(x)$ is a rational function that has no poles on the real axis and the degree of the polynomial in the denominator is at least one greater than that of the polynomial in the numerator.

Solution: Let $I = I_1 + iI_2 = \int_{-\infty}^{\infty} R(x)\ (\cos mx + i \sin mx)dx = \int_{-\infty}^{\infty} R(x)e^{imx}dx$ $m > 0$ and consider the function $R(z)e^{imz}$. Let $\sum r$ be the sum of the residues of $R(z)$ e^{imz} in the upper half plane. Then

$$I = \int_{-\infty}^{\infty} R(x)e^{imx}dx = 2\pi i \sum r$$

I_1 and I_2 are, respectively, the real and imaginary parts of I. In other words

$$\int_{-\infty}^{\infty} R(x)\cos mxdx = \mathrm{Re}\left\{2\pi i \sum r\right\}$$

$$\int_{-\infty}^{\infty} R(x)\sin mxdx = \mathrm{Im}\left\{2\pi i \sum r\right\}$$

Example 1: Evaluate $\int_{-\infty}^{\infty} \frac{\cos mx}{1+x^2}dx$

Solution: Consider the related function

$$\frac{e^{imz}}{1+z^2}$$

Its only pole in the upper half plane is $z = i$, and its residue that exists is

Residue at $z = i = \lim_{z \to i}\left[(z-i)\frac{e^{imz}}{(z-i)(z+i)}\right] = \frac{e^{-m}}{2i}$

So

$$\int_{-\infty}^{\infty} \frac{\cos mx}{1+x^2}dx = \mathrm{Re}\left\{2\pi i \frac{e^{-m}}{2i}\right\} = \pi e^{-m}$$

Type IV: The integral $I_1 = \int_{-\infty}^{\infty} f(x)\cos mxdx$ or $I_2 = \int_{-\infty}^{\infty} f(x)\sin mxdx$, where the associated complex function $f(z)$ is a meromorphic function which may have simple poles on the real axis and which approaches zero uniformly on any circle arc centered at $z = 0$ as the radius of the arc approaches infinity. See Fig. D.23

Let a function $f(z)$ satisfy the inequality $|f(z)| < K_\rho$ when z is on a circle arc Γ of radius ρ, and let K_ρ depend only on ρ so that the inequality holds for all z on Γ_ρ, regardless of the argument of z. If $K_\rho \to 0$ as $\rho \to \infty$, then $f(z)$ approaches zero uniformly on Γ_ρ as $\rho \to \infty$.

If ρ is allowed to become sufficiently large all poles in the upper half plane will fall within the contour shown in Fig. D.23.

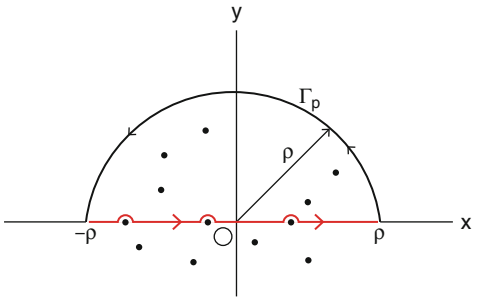


Fig. D.23 Contour Configuration

Solution: Consider the associated function $f(z)e^{imz} = f(z)\cos mx + f(z)\sin mx$. Let $\sum r$ be the sum of the residues of $f(z)e^{imz}$ at all poles lying in the upper half plane (not including those on the real axis). Let $\sum r'$ be the sum of the residues of $f(z)e^{imz}$ at all simple poles lying on the real axis. Then

$$PV\int_{-\infty}^{\infty} f(x)e^{imx}dx = 2\pi i\left(\sum r + \frac{1}{2}\sum r'\right)$$

$$PV\int_{-\infty}^{\infty} f(x)\cos dx = \mathrm{Re}\left\{2\pi i\left(\sum r + \frac{1}{2}\sum r'\right)\right\}$$

$$PV\int_{-\infty}^{\infty} f(x)\sin dx = \mathrm{Im}\left\{2\pi i\left(\sum r + \frac{1}{2}\sum r'\right)\right\}$$

Type V: The integral $\int_0^{\infty} x^{a-1}Q(x)dx$, where $Q(x)$ is analytic everywhere in the z plane except at a finite number of poles, none of which lies on the positive half of the real axis.

Solution: The solution is given by the following theorem

Theorem: *Let* Q(z) *be analytic everywhere in the* z *plane except at a finite number of poles, none of which lies on the positive half of the real axis. If* $|z^a Q(z)|$ *converges uniformly to zero when* $z \to 0$ *and when* $z \to \infty$, *then*

$$\int_0^{\infty} x^{a-1}Q(x)dx = \frac{\pi}{\sin a\pi}\sum \text{residues of } (-z)^{a-1}Q(z) \text{ at all its poles}$$

provided arg z *is taken in the interval* $(-\pi, \pi)$.

It should be noted that unless a is an integer, $(-z)^{a-1}$ *is a multiple-valued function which, using the defining formula* $a^z = e^{z\ln a}$, *is given by*

$$(-z)^{a-1} = e^{(a-1)\ln(-z)} = e(a-1)^{\left[\ln\left|z+i\arg(-z)\right]\right|} \quad -\pi < \arg z < \pi$$

Type VI: The integral $\int_0^\infty x^{-k}R(x)dx$ where $R(x)$ is a rational function of z which has no poles at $z = 0$ nor on the positive part of the real axis and k is not an integer.

Solution: To insure convergence of this integral it is necessary that it have the proper behavior at zero and infinity. It is sufficient that $\lim_{x\to 0} x^{1-k}R(x) = 0$ and $\lim_{x\to\infty} x^{1-k}R(x) = 0$. In evaluating the integral we employ the related function $z^{-k}R(z)$ which is a multiple-valued function. The branch of this function that is used is $z^{-k} = e^{-k(\ln|z|+i\arg z)}$.

Let $\sum r$ be the sum of the residues of $z^{-k}R(z)$ at the poles of $R(z)$. The integral is evaluated by the formula.

$$\int_0^\infty x^{-k}R(x)dx = \frac{\pi e^{i\pi k}}{\sin \pi k}\sum r$$

provided $z^{-k} = e^{-k(\ln|z|+i\arg z)}$

For types of integral not covered above, evaluation by the method of residues, when possible at all, usually requires considerable ingenuity in selecting the appropriate contour and in eliminating the integrals over all but the selected portion of the contour.

D.10.5 Few Solved Problems of Residue Integration

Example 1: If $|F(z)| \le \frac{M}{R^k}$ for $z = Re^{i\theta}$ where $k > 0$ and M is constant, prove that

$$\lim_{R\to\infty} \int_\Gamma e^{imz}F(z)\,dz = 0$$

Where Γ is the semicircle arc Fig. D.24 and m is a positive constant.

Solution: If $z = Re^{i\theta}$, and $\int_\Gamma e^{imz}F(z)dz = \int_0^\pi e^{imRe^{i\theta}}F\left(Re^{i\theta}\right)iRe^{i\theta}d\theta$ Then we have the following,

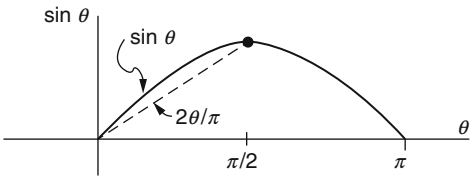


Fig. D.24 Contour Configuration

$$\left|\int_0^\pi e^{imRe^{i\theta}} F\left(Re^{i\theta}\right) iRe^{i\theta} d\theta\right| \le \int_0^\pi \left|e^{imRe^{i\theta}} F\left(Re^{i\theta}\right) iRe^{i\theta}\right| d\theta$$

$$= \int_0^\pi \left|e^{imR\cos\theta - mR\sin\theta} F\left(Re^{i\theta}\right) iRe^{i\theta}\right| d\theta$$

$$= \int_0^\pi e^{-mR\sin\theta} \left|F\left(Re^{i\theta}\right) iRe^{i\theta}\right| d\theta$$

$$\le \frac{M}{R^{k-1}} \int_0^\pi e^{-mR\sin\theta} d\theta = \frac{2M}{R^{k-1}} \int_0^{\pi/2} e^{-mR\sin\theta} d\theta$$

Now $\sin\theta \ge 2\theta/\pi$ for $0 \le \theta \le \pi/2$, as can be seen geometrically from Fig. D.59 or analytically by considering the derivative of $(\sin\theta)/\theta$, showing that it is a decreasing function.

As $R \to \infty$ this approaches zero, since m and k are positive, and the required result is proved.

Example 2: Show that $\displaystyle\int_0^\infty \frac{\cos mx}{x^2+1} dx = \frac{\pi}{2} e^{-m} \quad m > 0$.

Solution: Consider $\displaystyle\oint_C \frac{e^{imz}}{z^2+1} dz$ where C is the contour Fig. D.25. The integrand has simple poles at $z = \pm i$, but only $z = i$ lies inside C.

Residue at $z = i$ is $\lim_{z\to i} \left\{(z-i)\frac{e^{imz}}{(z-i)(z+i)}\right\} = \frac{e^{-m}}{2i}$.

Then

$$\oint_C \frac{e^{imz}}{z^2+1} dz = 2\pi i \left(\frac{e^{-m}}{2i}\right) = \pi e^{-m}$$

or

$$\int_{-R}^{R} \frac{e^{imx}}{x^2+1} dx + \int_\Gamma \frac{e^{imz}}{z^2+1} dz = \pi e^{-m}$$

i.e.,

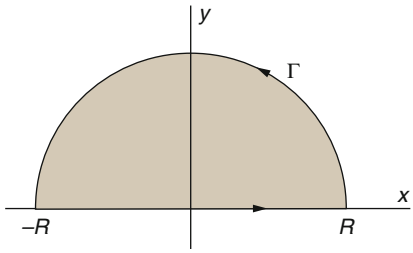


Fig. D.25 Contour Configuration

$$\int_{-R}^{R} \frac{\cos mx}{x^2+1}dx + i\int_{-R}^{R} \frac{\sin mx}{x^2+1}dx + \int_{\Gamma} \frac{e^{imz}}{z^2+1}dz = \pi e^{-m}$$

and so

$$2\int_{0}^{R} \frac{\cos mx}{x^2+1}dx + \int_{\Gamma} \frac{e^{imz}}{z^2+1}dz = \pi e^{-m}$$

Taking the limit as $R \to \infty$ and using Example 1 to show that the integral around Γ approaches zero, we obtain the required result.

Example 3: Show that $\int_0^{\infty} \frac{\sin x}{x}dx = \frac{\pi}{2}$.

Solution: The method of Example 3 in above leads us to consider the integral of e^{iz}/z around the contour of Fig. D.26. However, since $z = 0$ lies on this path of integration and since we cannot integrate through a singularity, we modify that contour by indenting the path at $z = 0$, as shown in Fig. D.26 below, which we call contour C' or *ABDEFGHJA*.

Since $z = 0$ is outside C', we have

$$\oint_{C'} \frac{e^{iz}}{z}dz = 0$$

or

$$\int_{-R}^{-\varepsilon} \frac{e^{ix}}{x}dx + \int_{HJA} \frac{e^{iz}}{z} + \int_{\varepsilon}^{R} \frac{e^{ix}}{x}dx + \int_{BDEFG} \frac{e^{iz}}{z}dz = 0$$

Replacing x by $-x$ in the first integral and combining with the third integral, we find

$$\int_{\varepsilon}^{R} \frac{e^{ix} - e^{-ix}}{x}dx + \int_{HJA} \frac{e^{iz}}{z}dz + \int_{BDEFG} \frac{e^{iz}}{z}dz = 0$$

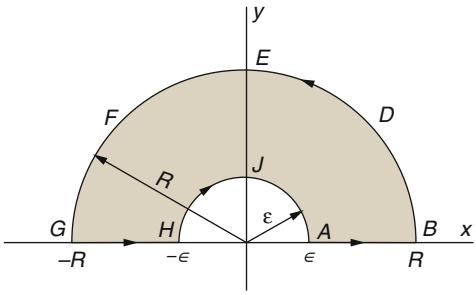


Fig. D.26 Contour Configuration

or

$$2i\int_{\varepsilon}^{R}\frac{\sin x}{x}dx=-\int_{HJA}\frac{e^{iz}}{z}dz-\int_{BDEFG}\frac{e^{iz}}{z}dz=0$$

Let $\varepsilon\to 0$ and $R\to\infty$. By Example 1 above, the second integral on the right approaches zero. Letting $z=\varepsilon e^{i\theta}$ in the first integral on the right hand side, we see that it approaches

$$-\lim_{\varepsilon\to 0}2i\int_{\pi}^{0}\frac{e^{i\varepsilon e^{i\theta}}}{\varepsilon e^{i\theta}}i\varepsilon e^{i\theta}d\theta=-\lim_{\varepsilon\to 0}\int_{\pi}^{0}ie^{i\,\varepsilon e^{i\theta}}d\theta=\pi i$$

since the limit can be taken under the integral sign.

Then we have

$$\lim_{\substack{R\to\infty\\ \varepsilon\to 0}}2i\int_{\varepsilon}^{R}\frac{\sin x}{x}dx=\pi i\quad\text{or}\quad\int_{\varepsilon}^{\infty}\frac{\sin x}{x}dx=\frac{\pi}{2}$$

Example 4: Show that $\displaystyle\int_0^\infty\frac{x^{p-1}}{1+x}dx=\frac{\pi}{\sin p\pi}0<p<1$

Solution: Consider $\displaystyle\oint_C\frac{z^{p-1}}{1+z}dz$. Since $z=0$ is a branch point, choose C as the contour of Fig. D.27 where the positive real axis is the branch line and where AB and GH are actually coincident with the x axis but are shown separated for visual purposes.

The integrand has the simple pole $z=-1$ inside C.

Residue at $z=-1=e^{i\pi}$ is $\lim_{z\to -1}(z+1)\frac{z^{p-1}}{1+z}=(e^{i\pi})^{p-1}=e^{(p-1)i\pi}$. Then $\displaystyle\oint_C\frac{z^{p-1}}{1+z}dz=2\pi ie^{(p-1)i\pi}$ or, omitting the integrand, then we

$$\int_{AB}+\int_{BDEFG}+\int_{GH}+\int_{HJA}=2\pi i e^{(p-1)i\pi}$$

We thus can write

$$\int_{\varepsilon}^{R}\frac{x^{p-1}}{1+x}dx+\int_{0}^{R}\frac{\left(Re^{i\theta}\right)^{p-1}iRe^{i\theta}}{1+Re^{i\theta}}d\theta+\int_{R}^{\varepsilon}\frac{\left(xe^{2i\pi}\right)^{p-1}}{1+xe^{2i\pi}}dx+\int_{R}^{0}\frac{\left(\varepsilon e^{i\theta}\right)^{p-1}i\varepsilon e^{i\theta}}{1+\varepsilon e^{i\theta}}d\theta$$
$$=2\pi i e^{(p-1)i\pi}$$

Where we have used $z=xe^{2\pi i}$ for the integral along GH, since the argument of z is increased by 2π in going around the circle $BDEFG$.

Taking the limit as $\varepsilon\to 0$ and $R\to\infty$, then noting that the second and fourth integrals approach zero, we find

$$\int_{0}^{\infty}\frac{x^{p-1}}{1+x}dx+\int_{\infty}^{0}\frac{e^{2i\pi(p-1)}x^{p-1}}{1+x}dx=2\pi e^{(p-1)i\pi}$$

or

$$\left(1-e^{2i\pi(p-1)}\right)\int_{0}^{\infty}\frac{x^{p-1}}{1+x}dx=2\pi i\theta e^{(p-1)i\pi}$$

so that

$$\int_{0}^{\infty}\frac{x^{p-1}}{1+x}dx=\frac{2\pi i\theta e^{(p-1)i\pi}}{1-e^{2i\pi(p-1)}}=\frac{2\pi i}{e^{p\pi i}-e^{-p\pi i}}=\frac{\pi}{\sin p\pi}$$

Example 5: Prove that $\Gamma(m)=2\int_{0}^{\infty}x^{2m-1}e^{-x^2}dx\quad m>0$

Solution: If $t=x^2$, we have $\Gamma(m)=2\int_{0}^{\infty}t^{m-1}e^{-t}dt=2\int_{0}^{\infty}\left(x^2\right)^{m-1}e^{-x^2}2xdx=2\int_{0}^{\infty}x^{2m-1}e^{-x^2}dx$

The result also holds if $\mathrm{Re}\{m\}>0$.

Example 6: Prove that $\Gamma(z)\Gamma(1-z)=\frac{\pi}{\sin\pi z}$

Solution: We first prove that for real values of z such that $0<z<1$. By analytic continuation we can then extend it to other values of z.

From Example 6, we have for $0<m<1$,

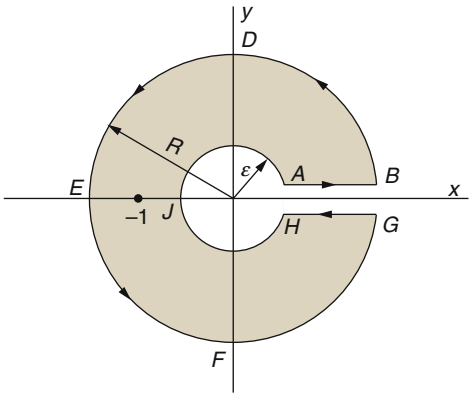


Fig. D.27 Contour Configuration

$$\Gamma(m)\Gamma(1-m) = \left\{2\int_0^\infty x^{2m-1}e^{-x^2}dx\right\}\left\{2\int_0^\infty y^{1-2m}e^{-y^2}dy\right\}$$

$$= 4\int_0^\infty\int_0^\infty x^{2m-1}y^{1-2m}e^{-\left(x^2+y^2\right)}dxdy$$

In terms of polar coordinate (r,θ) with $x = r\sin\theta$ this becomes

$$4\int_{\theta=0}^{\pi/2}\int_{r=0}^{\infty}\left(\tan^{1-2m}\theta\right)\left(re^{-r2}\right)drd\theta = 2\int_0^{\pi/2}\tan^{1-2m}\theta d\theta = \frac{\pi}{\sin m\pi}$$

See Example 4 in above with $x = \tan^2\theta$ and $p = 1-m$

Example 7: Prove that $\Gamma(z+1) = z\Gamma(z)$ and using the definition that for $\mathrm{Re}\{z\} > 0$, we define the gamma function by $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$.

Solution: Integrating by parts, we have if $\mathrm{Re}\{z\} > 0$,

$$\Gamma(z+1) = \int_0^\infty t^z e^{-t}dt = \lim_{M\to\infty}\int_0^M t^z e^{-t}dt$$

$$= \lim_{M\to\infty}\left\{(t^z)(-e^{-t})\big|_0^M - \int_0^M\left(zt^{z-1}\right)(-e^{-t})dt\right\}$$

$$= z\int_0^\infty t^{z-1}e^{-t}dt = z\Gamma(z)$$

Example 8: Prove that. $\Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-u^2}du = \sqrt{\pi}$

Solution: From Example 5 in above, letting $m = \frac{1}{2}$, we have

$$\Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-x^2}dx$$

From Example 6 in above, letting

$$\left\{\Gamma\left(\tfrac{1}{2}\right)\right\}^2 = \pi \quad \text{or} \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Since $\Gamma\left(\frac{1}{2}\right) > 0$. Thus the required result follows.

Another method: As in Example 6, we have

$$\left\{\Gamma\left(\tfrac{1}{2}\right)^2\right\} = \left\{2\int_0^\infty e^{-x^2}dx\right\}\left\{2\int_0^\infty e^{-y^2}dy\right\}$$

$$= 4\int_0^\infty\int_0^\infty e^{-\left(x^2+y^2\right)}dxdy = 4\int_{\theta=0}^{\pi/2}\int_{r=0}^\infty re^{-r^2}drd\theta = \pi$$

From which $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Example 9: Prove that $\int_0^\infty \sin x^2 dx = \int_0^\infty \cos x^2 dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}$.

Solution: The $\oint_C e^{iz^2}dz = 0$ or we can write (See Fig. D.28):

$$\int_{OA} e^{iz^2}dz + \int_{AB} e^{iz^2}dz + \int_{BO} e^{iz^2}dz = 0 \tag{D.46}$$

Now on OA, $z = x$ (from $x = 0$ to $x = R$); on AB, $z = Re^{i\theta}$ (from $\theta = 0$ to $\theta = \pi/4$); on BO,

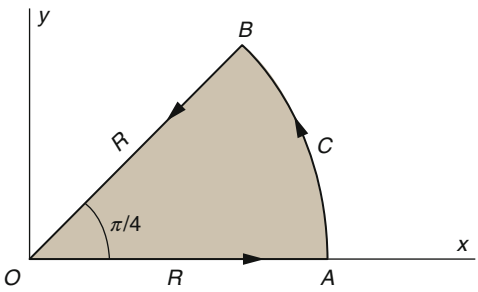


Fig. D.28 Contour Configuration

$$\int_0^R e^{ix^2}dx + \int_0^{\pi/4} e^{iRe^{i\theta}e^{2i\theta}} iRe^{i\theta}d\theta + \int_R^0 e^{ir^2}e^{i\pi/2}e^{i\pi/4}dr = 0 \tag{D.47}$$

i.e.

$$\int_0^R \left(\cos x^2 + \sin x^2\right)dx = e^{i\pi/4}\int_0^R e^{-r^2}dr - \int_0^{\pi/4} e^{iR^2\cos 2\theta - R^2\sin 2\theta} iRe^{i\theta}d\theta \tag{D.48}$$

We consider the limit of Eq. (D.48) as $R \to \infty$. The first integral on the right becomes (See Example 7 below),

$$e^{i\pi/4}\int_0^\infty e^{-r^2}dr = \frac{\sqrt{\pi}}{2}e^{i\pi/4} = \frac{1}{2}\sqrt{\frac{\pi}{2}} + \frac{i}{2}\sqrt{\frac{\pi}{2}}$$

The absolute value of the second integral on the right of Eq. (D.48) is

$$\begin{aligned}\left|\int_0^{\pi/4} e^{iR^2\cos 2\theta - R^2\sin 2\theta} iRe^{i\theta}d\theta\right| &\le \int_0^{\pi/4} e^{-R^2\sin 2\theta} iRe^{i\theta}d\theta \\ &= \int_0^{\pi/4} e^{-R^2\sin 2\theta}d\theta \\ &\le \frac{R}{2}\int_0^{\pi//2} e^{-R^2\phi/\pi}d\phi \\ &= \frac{\pi}{4R}\left(1 - e^{-R^2}\right)\end{aligned}$$

Where we have used the transformation $2\theta = \phi$ and the inequality $\sin\phi \ge 2\phi/\pi$, $0 \le \phi \le \pi/2$ (See Example 1 above). This shows that $R \to \infty$ the second integral on the right of Eq. (D.48) approaches zero. Then Eq. (D.48) becomes

$$\int_0^\infty \left(\cos x^2 + i\sin x^2\right)dx = \frac{1}{2}\sqrt{\frac{\pi}{2}} + i\sqrt{\frac{\pi}{2}}$$

and so equating real and imaginary parts we have, as required

$$\int_0^\infty \cos x^2 dx = \int_0^\infty \sin x^2 dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}$$

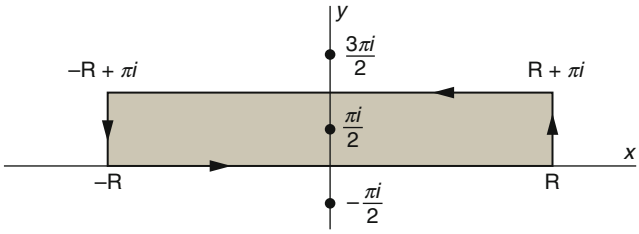


Fig. D.29 Contour Configuration

Example 10: Prove that. $\int_0^\infty \frac{\cosh ax}{\cosh x} = \frac{\pi}{2\cos(\pi a/2)}$ where $|a| < 1$

Solution: Consider $\oint_C \frac{e^{az}}{\cosh z} dz$ where C is a rectangle having vertices at $-R$, R, $R + \pi i$ and $-R + \pi i$ (see Fig. D.29).

The poles of $e^{az/\cosh z}$ are and occur where $\cosh z = 0$, i.e. $z = \left(n + \frac{1}{2}\right)\pi i$, $n = 0, \pm 1, \pm, \ldots$. The only pole enclosed by C is $\pi i/2$.

Residue of $\frac{e^{az}}{\cosh z}$ at $z = \pi i/2$ is

$$\lim_{z \to \pi i/2} (z - \pi i/2) \frac{e^{az}}{\cosh z} = \frac{e^{a\pi i/2}}{\sinh(\pi i/2)} = i \frac{e^{a\pi i/2}}{\sin(\pi/2)} = -ie^{a\pi i/2}$$

Then by residue theorem,

$$\oint_C \frac{e^{az}}{\cosh z} dz = 2\pi i\left(-ie^{a\pi i/2}\right) = 2\pi e^{a\pi i/2}$$

This can be written

$$\begin{aligned} \int_{-R}^{+R} \frac{e^{ax}}{\cosh x} dx + \int_0^\pi \frac{e^{a(R+iy)}}{\cosh(R+iy)} idy + \int_{+R}^{-R} \frac{e^{a(x+\pi i)}}{\cosh(x+\pi i)} dx \\ + \int_0^\pi \frac{e^{a(R+iy)}}{\cosh(R+iy)} idy = 2\pi e^{a\pi i/2} \end{aligned} \tag{D.49}$$

as $R \to \infty$ the second and fourth integrals on the left approach zero. To show this let us consider the second integral. Since

$$|\cosh(R+iy)| = \left|\frac{e^{R+iy} + e^{-R-iy}}{2}\right| \geq \frac{1}{2}\left\{\left|e^{R+iy}\right| - \left|e^{-R-iy}\right|\right\} = \frac{1}{2}\left(e^R - e^{-R}\right) \geq \frac{1}{4}e^R$$

We have

$$\left|\int_0^{\pi} \frac{e^{a(R+iy)}}{\cosh(R+iy)} idy\right| \leq \int_0^{\pi} \frac{e^{aR}}{\frac{1}{4}e^{R}} dy = 4\pi e^{(a-1)R}$$

and the result follows on nothing that the right side approaches zero as $R \to \infty$ since $|a| < 1$. In a similar manner we can show that the fourth integral on the left of Eq. (D.49) approaches zero as $R \to \infty$. Hence Eq. (D.49) becomes

$$\lim_{R\to\infty} \left\{ \int_{-R}^{+R} \frac{e^{ax}}{\cosh x} dx + e^{a\pi i/2} \int_{-R}^{+R} \frac{e^{ax}}{\cosh x} dx \right\} = 2\pi e^{a\pi i/2}$$

Since $\cosh(x + \pi i) = -\cosh x$. Thus

$$\lim_{R\to\infty} \int_{-R}^{+R} \frac{e^{ax}}{\cosh x} dx = \int_{-R}^{+R} \frac{e^{ax}}{\cosh x} dx = \frac{2\pi e^{a\pi i/2}}{1 + e^{a\pi i}} = \frac{2\pi}{e^{a\pi i/2} + e^{-a\pi i/2}} = \frac{\pi}{\cos(\pi a/2)}$$

Now

$$\int_{-\infty}^{0} \frac{e^{ax}}{\cosh x} dx + \int_{0}^{+\infty} \frac{e^{ax}}{\cosh x} dx = \frac{\pi}{\cos(\pi a/2)}$$

Then replacing x by $-x$ in the first integral, we have

$$\int_0^{+\infty} \frac{e^{-ax}}{\cosh x} dx + \int_0^{+\infty} \frac{e^{ax}}{\cosh x} dx = 2\int_0^{+\infty} \frac{e^{ax}}{\cosh x} dx = \frac{\pi}{\cos(\pi a/2)}$$

From which the required result follows.

Example 11: Prove that $\int_0^{\infty} \frac{\ln(x^2+1)}{x^2+1} dx = \pi \ln 2$

Solution: Consider $\oint_C \frac{\ln(z+i)}{z^2+1} dz$ around the contour C consisting of the real axis from $-R$ to R and the semicircle Γ of radius R (See Fig. D.30). The only pole of $\ln(z+i)/(z^2+1)$ inside C is the simple pole $z = i$, and the residue is

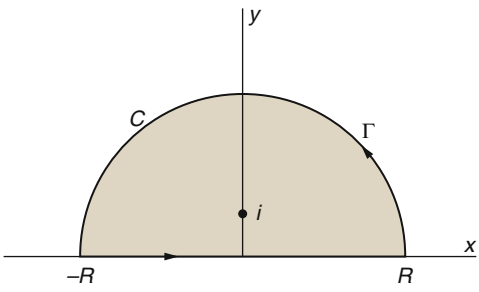


Fig. D.30 Contour Configuration

$$\lim_{z\to i}(z-i)\frac{\ln(z+i)}{(z-i)(z+i)}=\frac{\ln(2i)}{2i}$$

Hence by the residue theorem,

$$\oint_C \frac{\ln(z+i)}{z^2+1}dz=2\pi i\left\{\frac{\ln(2i)}{2i}\right\}=\pi\ln(2i)=\pi\ln 2+\frac{1}{2}\pi^{2i} \tag{D.50}$$

on writing $\ln(2i)=\ln 2+\ln i=\ln 2+\ln e^{\pi/2}=\ln 2+\pi i/2$ using principal values of the logarithm. The result can be written

$$\int_{-R}^{R}\frac{\ln(x+i)}{x^2+1}dx+\int_{\Gamma}\frac{\ln(z+i)}{z^2+1}dz=\pi\ln 2+\frac{1}{2}\pi^2 i$$

or

$$\int_{-R}^{0}\frac{\ln(x+i)}{x^2+1}dx+\int_{0}^{R}\frac{\ln(x+i)}{x^2+1}dx+\int_{\Gamma}\frac{\ln(z+i)}{z^2+1}dz=\pi\ln 2+\frac{1}{2}\pi^2 i$$

Replacing x by $-x$ in the first integral, this can be written

$$\int_{-R}^{0}\frac{\ln(i-x)}{x^2+1}dx+\int_{0}^{R}\frac{\ln(i+x)}{x^2+1}dx+\int_{\Gamma}\frac{\ln(z+i)}{z^2+1}dz=\pi\ln 2+\frac{1}{2}\pi^2 i$$

or, since $\ln(i-x)=\ln(i+x)=\ln\left(i^2-x^2\right)=\ln(x^2+1)+\pi i$

$$\int_{0}^{R}\frac{\ln(x^2+1)}{x^2+1}dx+\int_{0}^{R}\frac{\pi i}{x^2+1}dx+\int_{\Gamma}\frac{\ln(z+i)}{z^2+1}dz=\pi\ln 2+\frac{1}{2}\pi^2 i \tag{D.51}$$

As $R\to\infty$ we can show that the integral around Γ approaches zero. Hence on taking real parts we find, as required,

$$\lim_{R\to\infty}\int_{0}^{R}\frac{\ln(x^2+1)}{x^2+1}dx=\int_{0}^{\infty}\frac{\ln(x^2+1)}{x^2+1}dx=\pi\ln 2$$

Example 12: Prove that $\int_0^{\frac{\pi}{2}}\ln\sin x dx=\int_0^{\frac{\pi}{2}}\ln\cos x dx=-\frac{1}{2}\pi\ln 2$

Solution: Letting $x=\tan\theta$ in the result of Example 11, we find

$$\int_0^{\frac{\pi}{2}}\frac{\ln(\tan^2\theta+1)}{\tan^2\theta+1}\sec^2\theta d\theta=-2\int_0^{\frac{\pi}{2}}\ln\cos\theta d\theta=\pi\ln 2$$

from which

$$\int_0^{\frac{\pi}{2}} \ln\cos\theta d\theta = -\frac{1}{2}\pi\ln 2 \tag{D.52}$$

which establishes part of the required result. Letting $\theta = \pi/2 - \phi$ in Eq. (D.52), we find

$$\int_0^{\frac{\pi}{2}} \ln\sin x dx = -\frac{1}{2}\pi\ln 2$$

D.11 References

1. Brown JW, Churchill RV (2009) Complex variables and applications. 8th Edition, McGraw-Hill
2. Spiegel MR. Complex variables with an introduction to conformal mapping and its applications. Schaum's Outlines, McGraw-Hill
3. Euler's Formula and Taylor's Series - Wikipedia, the free encyclopedia

Appendix E
Short Course in Fourier and Laplace Transforms

It should be pointed out that both of these topics are far more in depth than what we'll be covering here. In fact you can do whole courses on each of these topics. What we'll be covering here are simply the basics of these topics that well need in order to do the work in dealing with our materials that were covered in Chaps. 5 and 6 particularly in order to be able to solve some of our boundary value problems. There are whole areas of both of these topics that we'll not be even touching on. What we'll be covering here are simply the basics of these topics that well need in order to do the work in those chapters. There are whole areas of both of these topics that we'll not be even touching on. The main point of this appendix is to get some of the basics out of the way that we'll need in those chapters where we'll take a look at one of the more common solution methods for differential and partial differential equations.

E.1 Fourier Transformation

In mathematics, the Fourier transform (often abbreviated FT) is an operation that transforms one complex-valued function of a real variable into another. In such applications as signal processing, the domain of the original function is typically in time and is accordingly called the time domain. The domain of the new function is typically called the frequency domain, and the new function itself is called the frequency domain representation of the original function. It describes which frequencies are present in the original function. This is analogous to describing a musical chord in terms of the individual notes being played. In effect, the Fourier transform decomposes a function into oscillatory functions. The term Fourier transforms refers both to the frequency domain representation of a function, and to the process or formula that "transforms" one function into the other.

The Fourier transform and its generalizations are the subject of Fourier analysis. In this specific case, both the time and frequency domains are unbounded linear

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

continua. It is possible to define the Fourier transform of a function of several variables, which is important for instance in the physical study of wave motion and optics. It is also possible to generalize the Fourier transform on discrete structures such as finite groups. The efficient computation of such structures, by fast Fourier transform, is essential for high-speed computing.

E.2 Definition

There are several common conventions for defining the Fourier transform of an integrable function $f : \mathbf{R} \to \mathbf{C}$ (Kaiser 1994) [5]. This article will use the definition:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \xi} dx \quad \text{for every real number } \xi.$$

When the independent variable x represents time (with SI unit of seconds), the transform variable ξ represents frequency (in hertz). Under suitable conditions, f can be reconstructed from $\hat{f}$ by the inverse transform:

$$\hat{f}(x) = \int_{-\infty}^{\infty} f(\xi)\, e^{-2\pi i x \xi} d\xi \quad \text{for every real number } x.$$

For other common conventions and notations, including using the angular frequency ω instead of the frequency ξ, see Other conventions and Other notations below. The Fourier transform on Euclidean space is treated separately, in which the variable x often represents position and ξ momentum.

The motivation for the Fourier transform comes from the study of Fourier series. In the study of Fourier series, complicated periodic functions are written as the sum of simple waves mathematically represented by Sines and Cosines. Due to the properties of sine and cosine it is possible to recover the amount of each wave in the sum by an integral. In many cases it is desirable to use Euler's formula, which states that $e2^{\pi i\theta} = \cos 2\pi\theta + i \sin 2\pi\theta$, to write Fourier series in terms of the basic waves $e2^{\pi i\theta}$. This has the advantage of simplifying many of the formulas involved and providing a formulation for Fourier series that more closely resembles the definition followed in this article. This passage from Sines and Cosines to complex exponentials makes it necessary for the Fourier coefficients to be complex valued. The usual interpretation of this complex number is that it gives either the amplitude (or size) of the wave present in the function and the phase (or the initial angle) of the wave. This passage also introduces the need for negative "frequencies". If θ were measured in seconds then the waves $e^{2\pi i\theta}$ and $e^{2-\pi i\theta}$ would both complete one cycle per

second, but they represent different frequencies in the Fourier transform. Hence, frequency no longer measures the number of cycles per unit time, but is closely related.

We may use Fourier series to motivate the Fourier transform as follows. Suppose that f is a function which is zero outside of some interval $[-L/2, L/2]$. Then for any $T \geq L$ we may expand f in a Fourier series on the interval $[-T/2, T/2]$, where the "amount" (denoted by c_n) of the wave $e^{2\pi inx/T}$ in the Fourier series of f is given by

$$\hat{f}\,(n/T) = c_n = \int_{-T/2}^{T/2} e^{-2\pi inx/T} f(x)dx$$

and f should be given by the formula

$$f(x) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \hat{f}\,(n/T)e^{2\pi inx/T}$$

If we let $\xi_n = n/T$, and we let $\Delta\xi = (N+1)/T - n/T = 1/T$, then this last sum becomes the Riemann sum

$$f(x) = \sum_{n=-\infty}^{\infty} \hat{f}\,(\xi_n)e^{2\pi ix\xi_n}\Delta\xi$$

By letting $T \to \infty$. This Riemann sum converges to the integral for the inverse Fourier transform given in the Definition section. Under suitable conditions this argument may be made precise (Stein and Shakarchi 2003) [6]. Hence, as in the case of Fourier series, the Fourier transform can be thought of as a function that measures how much of each individual frequency is present in our function, and we can recombine these waves by using an integral (or "continuous sum") to reproduce the original function.

The following images provide a visual illustration of how the Fourier transforms measures whether a frequency is present in a particular function. The function depicted $f(t) = \cos(6\pi t)e^{-\pi t^2}$ oscillates at 3 hertz (if t measures seconds) and tends quickly to 0. This function was specially chosen to have a real Fourier transform which can easily be plotted. The first image contains its graph. In order to calculate $\hat{f}\,(3)$ we must integrate $e^{-2\pi i(3t)}f(t)$. The second image shows the plot of the real and imaginary parts of this function. The real part of the integrand is almost always positive, this is because when $f(t)$ is negative, then the real part of $e^{-2\pi i(3t)}f(t)$ is negative as well. Because they oscillate at the same rate, when f(t) is positive, so is the real part of $e^{-2\pi i(3t)}$. The result is that when you integrate the real part of the integrand you get a relatively large number (in this case 0.5). On the other hand, when you try to measure a frequency that is not present, as in the case when we look at , the integrand oscillates enough so that the integral is very small. The general situation may be a bit more complicated than this, but this in spirit is how the

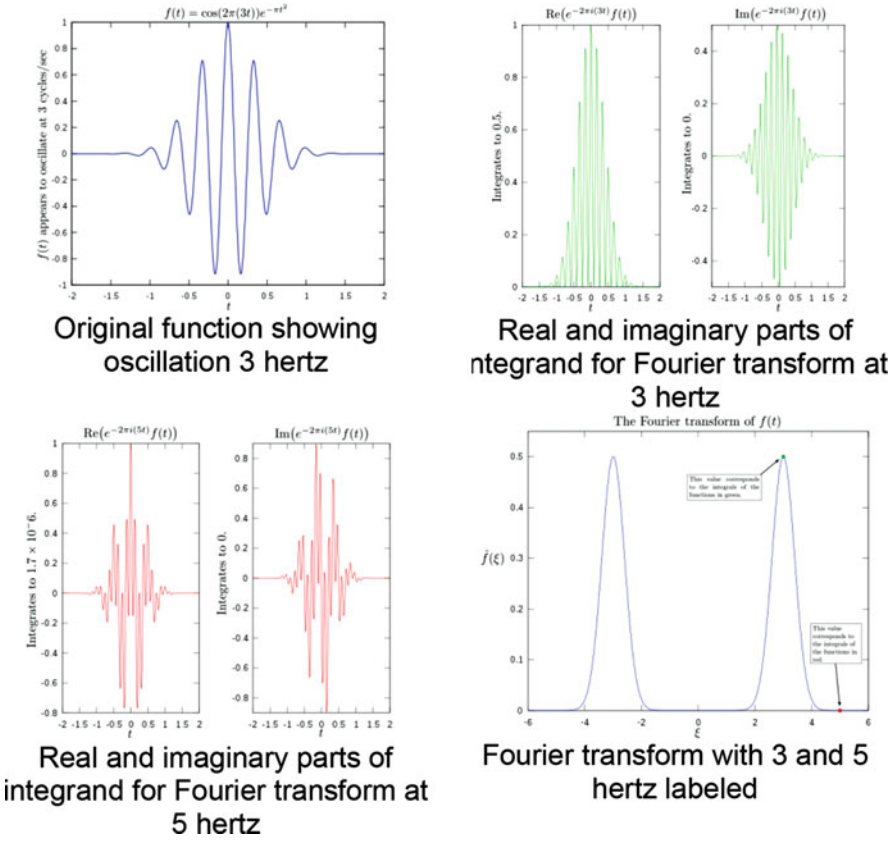

Original function showing oscillation 3 hertz
Real and imaginary parts of ntegrand for Fourier transform at 3 hertz
Real and imaginary parts of integrand for Fourier transform at 5 hertz
Fourier transform with 3 and 5 hertz labeled
The Fourier transform of f(t)

Fig. E.1 Fourier Analysis of $f(t) = \cos(6\pi t)e^{-\pi t^2}$

Fourier transform measures how much of an individual frequency is present in a function $f(t)$ (Fig. E.1).

E.3 Periodic Functions and Orthogonal Functions

This is going to be a short section. We just need to have a brief discussion about a couple of ideas that we'll be dealing with on occasion as we move into the next topic of this chapter.

Periodic Function

The first topic we need to discuss is that of a periodic function. A function is said to be **periodic** with **period** T if the following is true,

$$f(x+T) = f(x) \quad \text{for all } x$$

The following is a nice little fact about periodic functions.

Fact 1
If f and g are both periodic functions with period T then so is $f+g$ and fg.

This is easy enough to prove, so let's do that.

$$(f+g)(x+T) = f(x+T) = f(x) + g(x) = (f+g)(x)$$

$$(fg)(x+T) = f(x+T)g(x+T) = f(x)g(x) = (fg)(x)$$

The two periodic functions that most of us are familiar are sine and cosine and in fact we'll be using these two functions regularly in the remaining sections of this chapter. So, having said that let's close off this discussion of periodic functions with the following fact,

Fact 2
$\sin(\omega x)$ and $\cos(\omega x)$ are periodic functions with period $T = \frac{2\pi}{\omega}$.

E.4 Even and Odd Functions

The next quick idea that we need to discuss is that of even and odd functions.

Recall that a function is said to be **even** if,

$$f(-x) = -f(x)$$

and a function is said to be **odd** if,

$$f(-x) = -f(x)$$

The standard examples of even functions are $f(x) = x^2$ and $g(x) = \cos(x)$ while the standard examples of odd functions are $f(x) = x^3$ and $g(x) = \sin(x)$. The following fact about certain integrals of even/odd functions will be useful in some of our work.

The functions portrayed graphically in Figs. E.2 and E.3 below are odd and even respectively, but that of Fig. E.4 below is neither odd nor even.

In the Fourier series corresponding to an odd function, only sine terms can be present.

In the Fourier series corresponding to an even function, only cosine terms (and possibly a constant which we shall consider a cosine term) can be present.

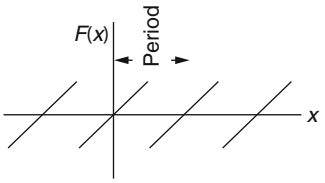


Fig. E.2 Different Functions Configurations

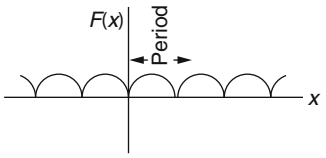


Fig. E.3 Different Functions Configurations

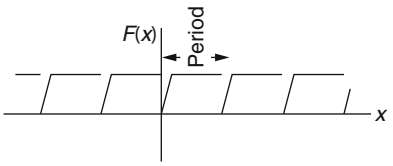


Fig. E.4 Different Functions Configurations

Fact 3
If $f(x)$ is an even function then,

$$\int_{-L}^{L} f(x)dx = 2\int_{0}^{L} f(x)dx$$

If $f(x)$ is an odd function then,

$$\int_{-L}^{L} f(x)dx = 0$$

Note that this fact is only valid on a "symmetric" interval, *i.e.* an interval in the form $[-L, L]$. If we are not integrating on a "symmetric" interval then the fact may or may not be true.

Example 1: If $f(x)$ is even, show that;

$$A_n = \frac{2}{L}\int_{0}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

$$B_n = 0$$

Solution: (a) From Fourier series we know that;

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

$$\frac{1}{L}\int_{-L}^{0} f(x)\cos\left(\frac{n\pi x}{L}\right)dx + \frac{1}{L}\int_{0}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

Letting $x = -u$,

$$\frac{1}{L}\int_{-L}^{0} f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

$$= -\frac{1}{L}\int_{-L}^{0} f(-u)\cos\left(\frac{-n\pi x}{L}\right)du$$

$$= \frac{1}{L}\int_{0}^{L} f(-u)\cos\left(\frac{n\pi u}{L}\right)du$$

$$= -\frac{1}{L}\int_{0}^{L} f(u)\cos\left(\frac{n\pi u}{L}\right)du$$

since by definition of an even function $f(-u) = f(u)$. Then by substituting for $x = -u$ again, we get the following result;

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

$$= \frac{1}{L}\int_{0}^{L} f(u)\cos\left(\frac{n\pi x}{L}\right)du + \frac{1}{L}\int_{0}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

$$= \frac{2}{L}\int_{0}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

(b) Again by Fourier series the coefficient B_n can be defined as follows:

$$B_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx$$

$$= \frac{1}{L}\int_{-L}^{0} f(x)\sin\left(\frac{n\pi x}{L}\right)dx + \frac{1}{L}\int_{0}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx \tag{E.1}$$

If we make the transformation of $x = -u$ in the first integral on the right hand side of Eq. (E.1), we obtain,

$$\begin{aligned}\frac{1}{L}\int_{-L}^{0} f(x)\sin\left(\frac{n\pi x}{L}\right)dx &= -\frac{1}{L}\int_{-L}^{0} f(-u)\sin\left(\frac{-n\pi u}{L}\right)du \\ &= -\frac{1}{L}\int_{0}^{L} f(-u)\sin\left(\frac{n\pi u}{L}\right)du \\ &= -\frac{1}{L}\int_{0}^{L} f(u)\sin\left(\frac{n\pi u}{L}\right)du \\ &= -\frac{1}{L}\int_{0}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx\end{aligned} \tag{E.2}$$

Where we have used the fact that for an even function $f(-u) = f(u)$ and in the last step that the dummy variable of integration u can be replaced by any other symbol, in particular x. Thus from Eq. (E.1), using Eq. (E.2), we have

$$\begin{aligned}B_n &= \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi u}{L}\right)dx \\ &= \frac{1}{L}\int_{-L}^{0} f(x)\sin\left(\frac{-n\pi x}{L}\right)dx + \frac{1}{L}\int_{0}^{L} f(x)\sin\left(\frac{nx\pi}{L}\right)dx \\ &= -\frac{1}{L}\int_{0}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx + \frac{1}{L}\int_{0}^{L} f(x)\sin\left(\frac{nx\pi}{L}\right)dx \\ &= 0\end{aligned}$$

E.5 Orthogonal Functions

The final topic that we need to discuss here is that of orthogonal functions. This idea will be integral to what we will be doing in the remainder of this chapter and in the next chapter as we discuss one of the basic solution methods for partial differential equations.

Let us first get the definition of orthogonal functions out of the way

Definition
Two non-zero functions, $f(x)$ and $g(x)$ are said to be **orthogonal on** $a \le x \le b$ set if ,

$$\int_a^b f(x)g(x)dx = 0$$

(continued)

A set of non-zero functions, $\{f_i(x)\}$ is said to be **mutually orthogonal** on $a \le x \le b$ (or just an **orthogonal set** if we are being lazy), $f_i(x)$ and $f_j(x)$ are orthogonal for every $a \le x \le b$. In other words,

$$\int_a^b f_i(x) f_j(x)\,dx = \begin{cases} 0 & i \ne j \\ c > 0 & i = j \end{cases}$$

Note that in the case of $i = j$ for the second definition we know that we will get a positive value from the integral because,

$$\int_a^b f_i(x) f_j(x)\,dx = \int_a^b \left[f_i(x)\right]^2 dx > 0$$

Recall that when we integrate a positive function we know the result will be positive as well.

Also note that the non-zero requirement is important because otherwise the integral would be trivially zero regardless of the other function we were using.

Before we work some examples there are a nice set of trig formulas that we'll need to help us with some of the integrals.

$$\sin\alpha\cos\beta = \frac{1}{2}[\sin(\alpha-\beta) + \sin(\alpha+\beta)]$$

$$\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta) + \cos(\alpha+\beta)]$$

$$\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha-\beta) + \cos(\alpha+\beta)]$$

Now let us work some examples that we will need over the course of the next couple of section

Example 1 Show that $\left\{\cos\left(\frac{n\pi x}{L}\right)\right\}_{n=0}^{\infty}$ is mutually orthogonal on $-L \le x \le L$.

Solution: This is not too difficult to do. All we really need to do is evaluate the following integral

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx$$

Before we start evaluating this integral let us notice that the integrand is the product of two even functions and so must also be even. This means that we can use Fact 3 above to write integral as,

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 2\int_{0}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx$$

There are two reasons for doing this. First having a limit of zero will often make the evaluation step a little easier and will be the case here. We will discuss the second reason after we are done with the example.

Now, in order to this integral we will actually need to consider three cases as below;

Case I: $n = m = 0$

In this case the integral is very easy and is

$$\int_{-L}^{L} dx = 2\int_{0}^{L} dx = 2L$$

Case II: $n = m \neq 0$

This integral is a little harder than the first case, but not by much (provided we recall a simple trigonometry formula). The integral for this case is

$$\int_{-L}^{L} \cos^2\left(\frac{n\pi x}{L}\right)dx = \int_{0}^{L} \cos\left(\frac{n\pi x}{L}\right)dx = \int_{0}^{L} 1 + \cos\left(\frac{n\pi x}{L}\right)dx$$

$$= \left(x + \frac{L}{2n\pi}\sin\left(\frac{2n\pi x}{L}\right)\right)\Bigg|_{0}^{L} = L + \frac{L}{2n\pi}\sin(2n\pi)$$

Now, at the point we need to recall that n is an integer and so $\sin(2n\pi) = 0$ and our final value for it is,

$$\int_{-L}^{L} \cos^2\left(\frac{n\pi x}{L}\right)dx = 2\int_{0}^{L} \cos^2\left(\frac{n\pi x}{L}\right)dx = L$$

The first two cases are really just showing that if $n = m$ the integral is not zero (as it should not be) and depending upon the n (and hence m) we get different values of the integral. Now we need to handle the third case scenario and this, in some ways, is the important case so farince.

Case III: $n = m \neq 0$

This is the "messiest" of the three that we have had to do here. Let us just start off by writing the integral down

$$\int_{-L}^{L} \cos^2\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 2\int_{0}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx$$

In this case we cannot combine/simplify as we did in the previous two cases. We can however, acknowledge that we have got a product of two cosines with different

argument and so we can use one of the trigonometry formulas above to break up the product as follows

$$\begin{aligned}\int_{-L}^{L}\cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx &= 2\int_{0}^{L}\cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx\\ &= \int_{0}^{L}\left[\cos\left(\frac{(n-m)\pi x}{L}\right)+\cos\left(\frac{(n+m)\pi x}{L}\right)\right]dx\\ &= \left[\frac{L}{(n-m)\pi}\sin\left(\frac{(n-m)\pi x}{L}\right)+\frac{L}{(n+m)\pi}\sin\left(\frac{(n+m)\pi x}{L}\right)\right]\\ &= \frac{L}{(n-m)\pi}\sin\left((n-m)\pi\right)+\frac{L}{(n+m)\pi}\sin\left((n+m)\pi\right)\end{aligned}$$

Now, we have shown if n and m are both integers and so $n-m$ and $n+m$ are also integers and so both of the sines above must be zero all together we get,

$$\int_{-L}^{L}\cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 2\int_{0}^{L}\cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 0$$

So, we have shown that if $n \neq m$ the integral is zero and if $n = m$ the value of the integral is a positive constant and so the set is mutually orthogonal.

In all of the work above we kept both forms of the integral at every step. By keeping both forms of the integral around we were able to show that not only is $\left\{\cos\left(\frac{n\pi x}{L}\right)\right\}_{n=0}^{\infty}$ mutually orthogonal on $-L \le x \le L$ but it is also mutually orthogonal on $0 \le x \le L$. The only difference is the value of the integral when $n = m$ and we can get those values from the work above.

Let us take a look at another example.

Example 2: Show that $\left\{\sin\left(\frac{n\pi x}{L}\right)\right\}_{n=1}^{\infty}$ is mutually orthogonal $-L \le x \le L$ and on $0 \le x \le L$.

Solution: First we will acknowledge from the start this time that we will be showing orthogonality on both of the intervals. Second, we need to start this set at $n = 1$ because if we used $n = 0$ the first function would be zero and we do not want the zero function to show on our list.

As with the first example all we really need to do is evaluate the following integral.

$$\int_{-L}^{L}\sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx$$

In this case integrand is the product of two odd functions and so must be even. This means that we can again use Fact 3 above to write the integral as,

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx = 2\int_{0}^{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx$$

We only have two cases to do for the integral here.

$$\underline{n = m}$$

Not much in case of this integral. It is pretty similar to the previous example second case

$$\begin{aligned}\int_{-L}^{L} \sin^2\left(\frac{n\pi x}{L}\right) dx &= 2\int_{0}^{L} \sin^2\left(\frac{n\pi x}{L}\right) dx = \int_{0}^{L} 1 - \cos\left(\frac{2n\pi x}{L}\right) \\ &= \left(x - \frac{L}{2n\pi}\sin\left(\frac{2n\pi x}{L}\right)\right)\Bigg|_{0}^{L} = L - \frac{L}{2n\pi}\sin(2n\pi) = L\end{aligned}$$

Summarizing up we get,

$$\int_{-L}^{L} \sin^2\left(\frac{n\pi x}{L}\right) dx = 2\int_{0}^{L} \sin^2\left(\frac{n\pi x}{L}\right) dx = L$$

$$\underline{n \neq m}$$

As with the previous example this can be a little messier but it is also nearly identical to the third case from the previous example so we will not show a lot of the work

$$\begin{aligned}\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx &= 2\int_{0}^{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx \\ &= \int_{0}^{L} \left[\cos\left(\frac{(n-m)\pi x}{L}\right) - \cos\left(\frac{(n-m)\pi x}{L}\right)\right] dx \\ &= \left[\frac{L}{(n-m)\pi}\sin\left(\frac{(n-m)\pi x}{L}\right) - \frac{L}{(n+m)}\sin\left(\frac{(n-m)\pi x}{L}\right)\right]\Bigg|_{0}^{L} \\ &= \frac{L}{(n-m)\pi}\sin((n-m)\pi) - \frac{1}{(n+m)\pi}\sin((n+m)\pi)\end{aligned}$$

As with the previous example we know that n and m are both integers and so both of the sines above must be zero and all together we get,

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx = 2\int_{0}^{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx = 0$$

So, we have shown that if $n \neq m$ the integral is zero and if $n = m$ the value of the integral is a positive constant and so the set is mutually orthogonal.

We have now shown that $\left\{ \sin\left(\frac{n\pi x}{L}\right) \right\}_{n=1}^{\infty}$ mutually orthogonal on $-L \leq x \leq L$ and on $-L \leq x \leq L$.

We need to work one more example in this section

Example 3: Show that $\left\{ \sin\left(\frac{n\pi x}{L}\right) \right\}_{n=1}^{\infty}$ and $\left\{ \cos\left(\frac{n\pi x}{L}\right) \right\}_{n=0}^{\infty}$ are mutually orthogonal on $-L \leq x \leq L$.

Solution: This example is a little different from the previous two examples. Here we want to show that together both sets are mutually orthogonal on $-L \leq x \leq L$. To show this we need to show three things. First (and second actually) we need to show that individually each set is mutually orthogonal and we have already done that in the previous two examples. The thirds (and only) thing we need to show here is that if we take one function from one set and another function from the other set and we integrate them we will get zero.

Also, note that this time we really do only want to do the one interval as the two sets, taken together, are not mutually orthogonal on $-L \leq x \leq L$. You might want to do the integral on this interval to verify that it will not always be zero.

So, let us take care of the one integral that we need to do here and then there is not a lot to do. Here is the integral

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right) dx$$

The integral in this case is the product of an off function (the sine) and an even function (the cosine) and so the integrand is an off function. Therefore, sine the integral is on a symmetric interval, *i.e.* $-L \leq x \leq L$ and so by Fact 3 above we know the integral must be zero or,

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right) dx = 0$$

So, in previous examples we have shown that on the interval $-L \leq x \leq L$ the two sets are mutually orthogonal individually and here we have shown that integrating a product of a sine and a cosine gives zero. Therefore, as a combined set they are also mutually orthogonal.

We have now worked three examples here dealing with orthogonalily and we should note that these were not just pulled out of the air as random examples to work. In the following sections (and Chaps. 4, 5 and 6) we will need the results from these examples. So, let us summarize those results up here.

Summary I

$\left\{\cos\left(\frac{n\pi x}{L}\right)\right\}_{n=0}^{\infty}$ and $\left\{\sin\left(\frac{n\pi x}{L}\right)\right\}_{n=1}^{\infty}$ are mutually orthogonal on $-L \le x \le L$ as individual sets and as a combined set.

$\left\{\cos\left(\frac{n\pi x}{L}\right)\right\}_{n=0}^{\infty}$ is mutually orthogonal on $0 \le x \le L$.

$\left\{\sin\left(\frac{n\pi x}{L}\right)\right\}_{n=1}^{\infty}$ is mutually orthogonal on $0 \le x \le L$.

We will also are going to need the results of the integrals themselves, both on $-L \le x \le L$ and on $0 \le x \le L$ so let us also summarize those up here as well so we can refer to them when we need to.

Summary II

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = \begin{cases} L & \text{if } n = m = 0 \\ \frac{L}{2} & \text{if } n = m \neq 0 \\ 0 & \text{if } n \neq m \end{cases}$$

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = \begin{cases} L & \text{if } n = m = 0 \\ \frac{L}{2} & \text{if } n = m \neq 0 \\ 0 & \text{if } n \neq m \end{cases}$$

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = \begin{cases} L & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx = \begin{cases} \frac{L}{2} & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 0$$

With this summary we will leave this section and move off into the second major topics of this appendix which is Fourier Series.

E.6 Fourier Sine Series

In this section we are going to start taking a look at Fourier series. We should point out that this is a subject that can span a whole class and what we'll be doing in this section (as well as the next couple of sections) is intended to be nothing more than a very brief look at the subject. The point here is to do just enough to allow us to do some basic solutions to partial differential equations in the next chapter. There are many topics in the study of Fourier series that we'll not even touch upon here.

So, with that out of the way let's get started, although we're not going to start off with Fourier series. Let's instead think back to our Calculus class where we looked at Taylor Series. With Taylor Series we wrote a series representation of a function $f(x)$, as a series whose terms were powers of $x - a$ for some $x = a$. With some conditions we were able to show that,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

and that the series will converge to $f(x)$ on $|x - a|$ for some R that will be dependent upon the function itself. There is nothing wrong with this, but it does require that derivatives of all orders exist at $x = a$. Or in other words $f^{(n)}(a)$ exists for $n = 0, 1, 2, \ldots$ Also for some functions the value of R may end up being quite small. These two issues (along with a couple of others) mean that this is not always the best way or writing a series representation for a function. In many cases it works fine and there will be no reason to need a different kind of series. There are times however where another type of series is either preferable or required.

We're going to build up an alternative series representation for a function over the course of the next couple of sections. The ultimate goal for the rest of this chapter will be to write down a series representation for a function in terms of sines and cosines.

We'll start things off by assuming that the function $f(x)$, we want to write a series representation for is an odd function (i.e. $f(-x) = -f(x)$). Because $f(x)$ is odd it makes some sense that should be able to write a series representation for this in terms of sines only (since they are also odd functions).

What we'll try to do here is write $f(x)$ as the following series representation, called a **Fourier Sine Series**, on $-L \le x \le L$.

$$\sum_{n=1}^{\infty} B_n \left(\frac{n\pi x}{L}\right)$$

There are a couple of issues to note here. First, at this point, we are going to assume that the series representation will converge to $f(x)$ on $-L \le x \le L$. We will be looking into whether or not it will actually converge in a later section. However, assuming that the series does converge to $f(x)$ it is interesting to note that, unlike

Taylor Series, this representation will always converge on the same interval and that the interval does not depend upon the function.

Second, the series representation will not involve powers of sine (again contrasting this with Taylor Series) but instead will involve sines with different arguments.

Finally, the argument of the sines $\frac{n\pi x}{L}$, may seem like an odd choice that was arbitrarily chosen and in some ways it was. For Fourier sine series the argument doesn't have to necessarily be this but there are several reasons for the choice here. First, this is the argument that will naturally arise in the next chapter when we use Fourier series (in general and not necessarily Fourier sine series) to help us solve some basic partial differential equations.

The next reason for using this argument is that fact that the set of functions that we chose to work with, $\left\{ \sin\left(\frac{n\pi x}{L}\right) \right\}_{n=1}^{\infty}$ in this case, need to be *orthogonal* on the given interval, $-L \le x \le L$ in this case, and note that in the last section we showed that in fact they are. In other words, the choice of functions we're going to be working with and the interval we're working on will be tied together in some way. We can use a different argument, but will need to also choose an interval on which we can prove that the sines (with the different argument) are orthogonal.

So, let's start off by assuming that given an odd function, $f(x)$, we can in fact find a Fourier sine series, of the form given above, to represent the function on $-L \le x \le L$. This means we will have,

$$f(x) = \sum_{n=1}^{\infty} B_n \left(\frac{n\pi x}{L}\right)$$

As noted above we'll discuss whether or not this even can be done and if the series representation does in fact converge to the function in later section. At this point we're simply going to assume that it can be done. The question now is how to determine the coefficients, B_n, in the series.

Let's start with the series above and multiply both sides by $\sin\left(\frac{m\pi x}{L}\right)$ where m is a fixed integer in the range $\{0, 1, 2, \ldots\}$. In other words we multiply both sides by any of the sines in the set of sines that we're working with here. Doing this gives,

$$f(x) \sin\left(\frac{m\pi x}{L}\right) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right)$$

Now, let's integrate both sides of this from $x - L$ to $x = L$.

$$\int_{-L}^{L} f(x) \sin\left(\frac{m\pi x}{L}\right) dx = \int_{-L}^{L} \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx$$

At this point we've got a small issue to deal with. We know from Calculus that an integral of a finite series (more commonly called a finite sum) is nothing more than

the (finite) sum of the integrals of the pieces. In other words for finite series we can interchange an integral and a series. For infinite series however, we cannot always do this. For some integrals of infinite series we cannot interchange an integral and a series. Luckily enough for us we actually can interchange the integral and the series in this case. Doing this step and factorig the constant gives, B_n, out of the integral gives,

$$\int_{-L}^{L} f(x)\sin\left(\frac{m\pi x}{L}\right)dx = \sum_{n=1}^{\infty}\int_{-L}^{L} B_n \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx$$

$$= \sum_{n=1}^{\infty} B_n \int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx$$

Now, recall from the last section we proved that $\left\{\sin\left(\frac{n\pi x}{L}\right)\right\}_{n=1}^{\infty}$ is orthogonal on $-L \le x \le L$ and that,

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx = \begin{cases} L & \text{if } n = m \\ 0 & \text{if } n \ne m \end{cases}$$

So, what does this mean for us. As we work through the various values of n in the series and compute the value of the integrals all but one of the integrals will be zero. The only non-zero integral will come when we have, $n = m$ in which case the integral has the value of L. Therefore, the only non-zero term in the series will come when we have $n = m$ and our equation becomes,

$$\int_{-L}^{L} f(x)\sin\left(\frac{m\pi x}{L}\right)dx = B_m L$$

Finally all we need to do is divide by L and we know have an equation for each of the coefficients.

$$B_m = \frac{1}{L}\int_{-L}^{L} \sin\left(\frac{m\pi x}{L}\right)dx \quad m = 1, 2, 3, \ldots$$

Next, note that because we're integrating two odd functions the integrand of this integral is even and so we also know that,

$$B_m = \frac{2}{L}\int_{0}^{L} \sin\left(\frac{m\pi x}{L}\right)dx \quad m = 1, 2, 3, \ldots$$

Summarizing all this work up the Fourier sine series of an odd function $f(x)$ on $-L \le x \le L$ is given by,

$$f(x) = \sum_{n=1}^{\infty} B_n \left(\frac{n\pi x}{L}\right) \qquad B_n = \frac{2}{L}\int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx \quad n = 1, 2, 3, \ldots$$
$$= \frac{2}{L}\int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx \qquad n = 1, 2, 3, \ldots$$

Let's take a quick look at an example.

Example 1: Find the Fourier sine series for $f(x) = x$ on $-L \le x \le L$.

Solution: First note that the function we're working with is in fact an odd function and so this is something we can do. There really isn't much to do here other than to compute the coefficients for $f(x) = x$.

Here is that work and note that we're going to leave the integration by parts details to you to verify. Don't forget that n, L, and π are constants!

$$B_n = \frac{2}{L}\int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx = \frac{2}{L}\left(\frac{L}{n^2\pi^2}\right) \quad \left(L \sin\left(\frac{n\pi x}{L}\right) - n\pi x \cos\left(\frac{n\pi x}{L}\right)\right)\Big|_0^L$$

$$\frac{2}{n^2\pi^2}(L \sin(n\pi) - n\pi L \cos(n\pi))$$

These integrals can, on occasion, be somewhat messy especially when we use a general L for the endpoints of the interval instead of a specific number.

Now, taking advantage of the fact that n is an integer we know that $\sin(n\pi) = 0$ and that $\cos(n\pi) = (-1)^n$. We therefore have,

$$B_n = \frac{2}{n^2\pi^2}(-n\pi L(-1)^n) = \frac{(-1)^{n+1}2L}{n\pi} \quad n = 1, 2, 3, \ldots$$

The Fourier sine series is then,

$$x = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}2L}{n\pi} \sin\left(\frac{n\pi x}{L}\right) = \frac{2L}{\pi}\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin\left(\frac{n\pi x}{L}\right)$$

At this point we should probably point out that we'll be doing most, if not all, of our work here on a general interval ($-L \le x \le L$ or $0 \le x \le L$) instead of intervals with specific numbers for the endpoints. There are a couple of reasons for this. First, it gives a much more general formula that will work for any interval of that form which is always nice. Secondly, when we run into this kind of work in the next chapter it will also be on general intervals so we may as well get used to them now.

Now, finding the Fourier sine series of an odd function is fine and good but what if, for some reason, we wanted to find the Fourier sine series for a function that is not odd? To see how to do this we're going to have to make a change. The above work was done on the interval $-L \le x \le L$. In the case of a function that is not odd

we'll be working on the interval $0 \le x \le L$. The reason for this will be made apparent in a bit.

So, we are now going to do is to try to find a series representation for $f(x)$ on the interval $0 \le x \le L$ that is in the form,

$$\sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

or in other words,

$$f(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

As we did with the Fourier sine series on $-L \le x \le L$ we are going to assume that the series will in fact converge to $f(x)$ and we'll hold off discussing the convergence of the series for a later section.

There are two methods of generating formulas for the coefficients, B_n, although we'll see in a bit that they really the same way, just looked at from different perspectives.

The first method is to just ignore the fact that $f(x)$ is odd and proceed in the same manner that we did above only this time we'll take advantage of the fact that we proved in the previous section that $\left\{ \sin\left(\frac{n\pi x}{L}\right) \right\}_{n=1}^{\infty}$ also forms an orthogonal set on $0 \le x \le L$ and that,

$$\int_0^L \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx = \begin{cases} \frac{0}{2} & \text{if } n = m \\ 0 & \text{if } n \ne m \end{cases}$$

So, if we do this then all we need to do is multiply both sides of our series by $\sin\left(\frac{m\pi x}{L}\right)$, integral from 0 to L and interchange the integral and series to get,

$$\int_0^L f(x) \sin\left(\frac{m\pi x}{L}\right) dx = \sum_{n=1}^{\infty} B_n \frac{2}{L} \int_0^L \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx$$

Now, plugging in for the integral we arrive at,

$$\int_0^L f(x) \sin\left(\frac{m\pi x}{L}\right) dx = B_m \left(\frac{L}{2}\right)$$

Upon solving for the coefficient we arrive at,

$$B_m = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{m\pi x}{L}\right)dx \quad \text{m} = 1, 2, 3, \ldots$$

Note that this is identical to the second form of the coefficients that we arrived at above by assuming $f(x)$ was odd and working on the interval $-L \le x \le L$. The fact that we arrived at essentially the same coefficients is not actually all the surprising as we'll see once we've looked the second method of generating the coefficients. Before we look at the second method of generating the coefficients we need to take a brief look at another concept. Given a function, $f(x)$, we define the **odd extension** of $f(x)$ to be the new function,

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le L \\ -f(-x) & \text{if } -L \le x \le L \end{cases}$$

It's pretty easy to see that this is an odd function.

$$g(-x) = -f(-(-x)) = f(x) = -g(x) \quad \text{for} \quad 0 < x < L$$

and we can also know that on $0 \le x \le L$ we have that $g(x) = f(x)$. Also note that if $f(x)$ is already an odd function then we in fact get $g(x) = f(x)$ on $-L \le x \le L$.

Let's take a quick look at a couple of odd extensions before we proceed any further.

Example 2: Sketch the odd extension of each of the given functions.

$$f(x) = L - x \quad \text{on} \quad 0 \le x \le L$$

$$f(x) = 1 + x^2 \quad \text{on} \quad 0 \le x \le L$$

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \le x \le \frac{L}{2} \\ x - \frac{L}{2} & \text{if} \frac{L}{2} \le x \le L \end{cases}$$

Solution: Not much to do with this other than to define the odd extension and then sketch it.

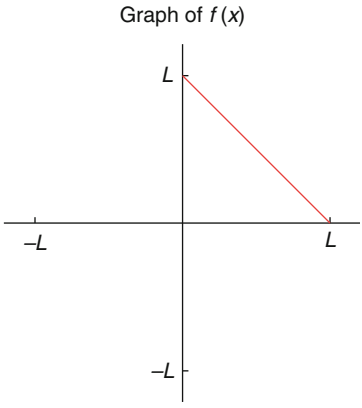
(a) $f(x) = L - x$ on $0 \le x \le L$

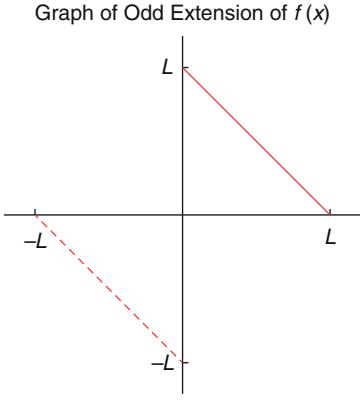
Here is the odd extension of this function.

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le \dfrac{L}{2} \\ -f(-x) & \text{if } -L \le x \le 0 \end{cases}$$

$$= \begin{cases} L - x & \text{if } 0 \le x \le L \\ -L - x & \text{if } -L \le x \le 0 \end{cases}$$

Below is the graph of both the function and its odd extension. Note that we've put the "extension" in with a dashed line to make it clear from the portion of the function that is being added to allow us to get the odd extension.





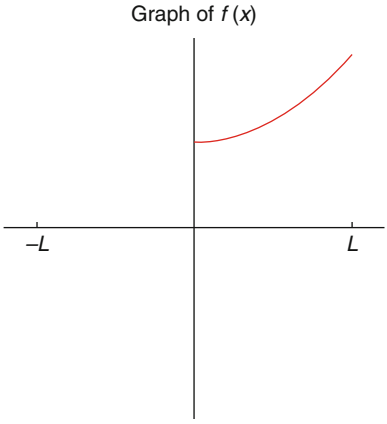
(b) $f(x) = 1 + x^2$ on $0 \le x \le L$

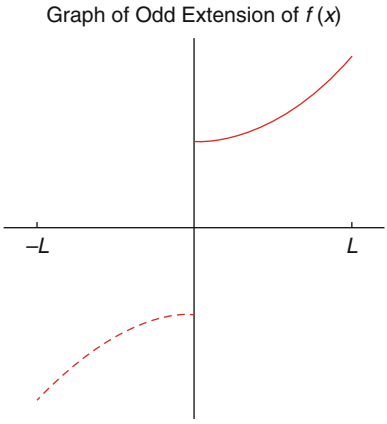
First note that this is clearly an even function. That does not however mean that we can't define the odd extension for it. The odd extension for this function is:

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le \dfrac{L}{2} \\ -f(-x) & \text{if } -L \le x \le 0 \end{cases}$$

$$= \begin{cases} 1 + x^2 & \text{if } \quad 0 \le x \le L \\ -1 - x^2 & \text{if } \quad -L \le x \le 0 \end{cases}$$

The sketch of the original function and its odd extension are:



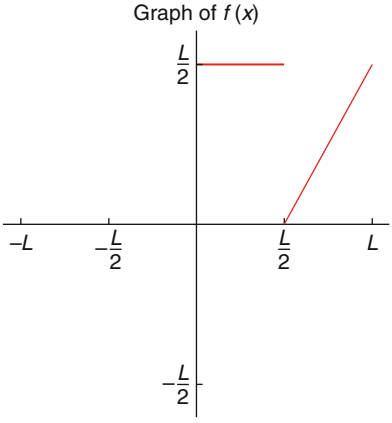


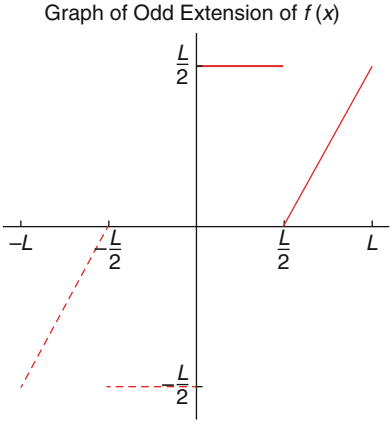
(c) $f(x) = \begin{cases} \dfrac{l}{2} & \text{if } \quad 0 \le x \le \dfrac{L}{2} \\ x - \dfrac{L}{2} & \text{if } \quad \dfrac{L}{2} \le x \le L \end{cases}$

Let's first write down the odd extension for this function.

$$g(x) = \begin{cases} f(x) & \text{if} \quad 0 \le x \le L \\ -f(-x) & \text{if} \quad -L \le x \le 0 \end{cases} = \begin{cases} x - \dfrac{L}{2} & \text{if} \quad \dfrac{L}{2} \le x \le L \\ \dfrac{L}{2} & \text{if} \quad 0 \le x \le \dfrac{L}{2} \\ -\dfrac{L}{2} & \text{if} \quad -\dfrac{L}{2} \le x \le \dfrac{L}{2} \\ -\dfrac{L}{2} & \text{if} \quad -L \le x \le -\dfrac{L}{2} \end{cases}$$

The sketch of the original function and its odd extension are:





With the definition of the odd extension (and a couple of examples) out of the way we can now take a look at the second method for getting formulas for the coefficients of the Fourier sine series for a function $f(x)$ on $0 \le x \le L$. First, given such a function define its odd extension as above. At this point, because $g(x)$ is an

odd function, we know that on $-L \le x \le L$ the Fourier sine series for $g(x)$ (and NOT $f(x)$ yet) is:

$$g(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) \quad B_n = \frac{2}{L}\int_0^L g(x)\sin\left(\frac{n\pi x}{L}\right)dx \quad n = 1, 2, 3, \ldots$$

However, because we know that $g(x) = f(x)$ on $0 \le x \le L$ we can also see that as long as we are on $0 \le x \le L$ we have:

$$f(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) \quad B_n = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right)dx \quad n = 1, 2, 3, \ldots$$

So, exactly the same formula for the coefficients regardless of how we arrived at the formula and the second method justifies why they are the same here as they were when we derived them for the Fourier sine series for an odd function.

Now, let's find the Fourier sine series for each of the functions that we looked at in Example 2.

Note that again we are working on general intervals here instead of specific numbers for the right endpoint to get a more general formula for any interval of this form and because again this is the kind of work we'll be doing in the next chapter.

Also, we'll again be leaving the actually integration details up to you to verify. In most cases it will involve some fairly simple integration by parts complicated by all the constants (n, L, π, *etc.*) that show up in the integral.

Example 3: Find the Fourier sine series for $f(x) = L - x$ on $-L \le x \le L$.

Solution: There really isn't much to do here other than computing the coefficients so here they are:

$$\begin{aligned} B_n &= \frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right) - dx = \frac{2}{L}\int_0^L (L-x)\sin\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{2}{L}\left(-\frac{L}{n^2\pi^2}\right)\left[L\sin\left(\frac{n\pi x}{L}\right) - n\pi(x-L)\cos\left(\frac{n\pi x}{L}\right)\right]\Bigg|_0^L \\ &= \frac{2}{L}\left[\frac{L^2}{n^2\pi^2}(n\pi - \sin(n\pi))\right] = \frac{2L}{n\pi} \end{aligned}$$

In the simplification process don't forget that n is an integer.

So, with the coefficients we get the following Fourier sine series for this function.

In the next example it is interesting to note that while we started out this section looking only at odd functions we're now going to be finding the Fourier sine series

of an even function on $0 \le x \le L$. Recall however that we're really finding the Fourier sine series of the odd extension of this function and so we're okay.

Example 4: Find the Fourier sine series for $f(x) = 1 + x^2$ on $0 \le x \le L$.

Solution: In this case the coefficients are liable to be somewhat messy given the fact that the integrals will involve integration by parts twice. Here is the work for the coefficients.

$$B_n = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right) - dx = \frac{2}{L}\int_0^L \left(L - x^2\right)\sin\left(\frac{n\pi x}{L}\right)dx$$

$$= \frac{2}{L}\left(\frac{L}{n^2\pi^2}\right)\left[2L^2 - n\pi\left(1 + x^2\right)\cos\left(\frac{n\pi x}{L}\right) + 2Ln\pi x\sin\left(\frac{n\pi x}{L}\right)\right]\Bigg|_0^L$$

$$= \frac{2}{L}\left(\frac{L^2}{n^2\pi^3}\right)\left[\left(2L^2 - n^2\pi^2\left(1 + L^2\right)\right)\cos(n\pi) + 2L^2 n\pi\sin(n\pi) - \left(2L^2 - n^2\pi^2\right)\right]$$

$$= \frac{2}{n^3\pi^3}\left[\left(2L^2 - n^2\pi^2\left(1 + L^2\right)\right)(-1)^n - 2L^2 + n^2\pi^2\right]$$

As noted above the coefficients are not the most pleasant ones, but there they are. The Fourier sine series for this function is then,

$$f(x) = \sum_{n=1}^{\infty}\frac{2}{n^3\pi^3}\left[\left(2L^2 - n^2\pi^2\left(1 + L^2\right)\right)(-1)^n - 2L^2 + n^2\pi^2\right]\sin\left(\frac{n\pi}{L}\right)$$

In the last example of this section we'll be finding the Fourier sine series of a piecewise function and this can definitely complicate the integrals a little, but they do show up on occasion and so we need to be able to deal with them.

Example 5: Find the Fourier sine series for $f(x) = \begin{cases} \dfrac{L}{2} & \text{if } 0 \le x \le L \\ x - \dfrac{L}{2} & \text{if } \dfrac{L}{2} \le x \le L \end{cases}$ on $0 \le x \le L$.

Solution: Here is the integral for the coefficients.

$$\int_0^{\frac{L}{2}} \sin\left(\frac{n\pi x}{L}\right) dx = -\left(\frac{L}{2}\right)\left(\frac{L}{n\pi}\right)\cos\left(\frac{n\pi x}{L}\right)\Big|_0^{\frac{L}{2}} = \frac{L^2}{2\pi n}\left(1 - \cos\left(\frac{n\pi}{2}\right)\right)$$

$$\int_{\frac{L}{2}}^{L} \left(x - \frac{L}{2}\right)\sin\left(\frac{n\pi x}{L}\right) dx \frac{L}{n^2\pi^2}\left[L\sin\left(\frac{n\pi x}{L}\right) - n\pi\left(x - \frac{L}{2}\right)\cos\left(\frac{n\pi x}{L}\right)\right]\Bigg|_{\frac{L}{2}}^{L}$$

$$= \frac{L}{n^2\pi^2}\left[L\sin(n\pi) - \frac{n\pi L}{2}\cos(n\pi) - L\sin\left(\frac{n\pi}{2}\right)\right]$$

$$= -\frac{L^2}{n^2\pi^2}\left[\frac{n\pi(-1)^n}{2} + \sin\left(\frac{n\pi}{2}\right)\right]$$

Note that we need to split the integral up because of the piecewise nature of the original function. Let's do the two integrals separately

Putting all of this together gives,

So, the Fourier sine series for this function is:

$$f(x) = \sum_{n=1}^{\infty} \frac{L}{n\pi}\left[1 + (-1)^{n+1} - \cos\left(\frac{n\pi}{2}\right) + \frac{2}{n\pi}\sin\left(\frac{n\pi}{2}\right)\right]\sin\left(\frac{n\pi x}{L}\right)$$

As the previous two examples have shown that the coefficients for these can be quite messy but that will often be the case and so we shouldn't let that get us too excited.

E.7 Fourier Cosine Series

In this section we are going to take a look at Fourier Cosine Series. We will start off much as we did in the previous section where we look at Fourier Sine Series. Let us start by assuming that the function, $f(x)$ that we will be working initially is an even function (i.e., $f(-x) = f(\lambda)$) and that we want to write a series representation for this function on $-L \le x \le L$ in terms of cosines (which are also even). In other words we are going to look for the following:

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right)$$

This series is called a **Fourier Cosine Series** and not that in this case (unlike with Fourier Sine Series) we are able to start the series representation at $n = 0$ since that term will not be zero as it was with sines. Also, as with Fourier Sine Series, the argument of $\frac{n\pi x}{L}$ in the cosines is being used only because it is the argument that we

will be running into in the next section. The only real requirement here is that the given set of functions we are using be orthogonal on the interval we are working on.

Note as well that we are assuming that the series will in fact converge to $f(x)$ on $-L \le x \le L$ at this point. In a later section we will be looking into the convergence of this series in more detail. So, to determine a formula for the coefficient, A_n, we will use the fact that $\left\{ \cos\left(\frac{n\pi x}{L}\right) \right\}_{n=0}^{\infty}$ do form an orthogonal set on the interval $-L \le x \le L$ as we showed in previous section. In that section we also derived the following formula that we will need in a bit.

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right) dx = \begin{cases} 2L & \text{if} \quad n = m = 0 \\ L & \text{if} \quad n = m \ne 0 \\ 0 & \text{if} \quad n \ne m \end{cases}$$

We will get a formula for the coefficient in almost exactly the same fashion that we did in the previous section. We will start with the representation above and multiply both sides by $\cos\left(\frac{n\pi x}{L}\right)$ where m is a fixed integer in the range $\{0, 1, 2, 3, \cdots\}$. Doing this gives:

$$f(x) \cos\left(\frac{m\pi x}{L}\right) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right)$$

Next, we integrate both sides from $x = -L$ to $x = L$ and as we were able to do with the Fourier Sine Series we can again interchange the integral and the series.

$$\begin{aligned} \int_{-L}^{L} f(x) \cos\left(\frac{m\pi x}{L}\right) dx &= \int_{-L}^{L} A_n \cos\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right) dx \\ &= \sum_{n=0}^{\infty} A_n \int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right) dx \end{aligned}$$

We now know that the all of the integrals on the right side will be zero except when because the set of cosines from orthogonal set on the interval $-L \le x \le L$. However, we need to be careful about the value of m (or n depending on the letter you want to use). So, after evaluating all of the integrals we arrive at the following set of formula for the coefficients.

$\underline{m = 0}$

$$\int_{-L}^{L} f(x)dx = A_0(2L) \quad \Rightarrow \quad A_0 = \left(\frac{1}{2L}\right)\int_{-L}^{L} f(x)dx$$

$\underline{m \neq 0}$

$$\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right)dx = A_m(L) \quad \Rightarrow \quad A_m = \left(\frac{1}{L}\right)\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right)dx$$

Summarizing everything up then, the Fourier Cosine Series of an even function, $f(x)$ on $-L \le x \le L$ is give by,

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) \quad A_n = \begin{cases} \left(\frac{1}{2L}\right)\int_{-L}^{L} f(x)dx & n = 0 \\ \left(\frac{1}{L}\right)\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx & n \neq 0 \end{cases}$$

Finally, before we work an example, let us notice that because both $f(x)$ and the cosines are even and the integrand in both of the integrals above is even and so we can write the formula for the A_n's as follows,

$$A_n = \begin{cases} \frac{1}{L}\int_{-L}^{L} f(x)dx & n = 0 \\ \frac{2}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx & n \neq 0 \end{cases}$$

Now let us take a look at an example.

Example 1: Find the Fourier cosine series of $f(x) = x^2$ on $-L \le x \le L$.

Solution: We clearly have an even function here and so all we really need to do is compute the coefficients and they are liable to be a little messy because we will need to do integration by parts twice. We will leave most of the actual integration details to you to verify.

$$A_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx = \frac{1}{L}\int_{0}^{L} x^2 dx = \frac{1}{L}\left(\frac{L^3}{3}\right) = \frac{L^3}{3}$$

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{2}{L}\int_{0}^{L} x^2\cos\left(\frac{n\pi x}{L}\right)dx$$

$$= \tfrac{2}{L}\left(\frac{L}{n^3\pi^3}\right)\left(2Ln\pi\cos\left(\frac{n\pi x}{L}\right) + \left(n\pi x^2 - 2L^2\right)\sin\left(\frac{n\pi x}{L}\right)\right)\Big|_0^L$$

$$= \frac{2}{n^3\pi^3}\left(2L^2 n\pi\cos(n\pi) + \left(n^2\pi^2L^2 - 2L^2\right)\sin(n\pi)\right)$$

$$= \frac{4L^2(-1)^n}{n^2\pi^2} \quad \text{for } n = 1, 2, 3, \ldots$$

The coefficients are then,

$$A_0\frac{L^2}{3}\text{and} \quad A_n = \frac{4L^2(-1)^n}{n^2\pi^2}\text{for } n = 1, 2, 3, \ldots$$

The Fourier Cosine Series is then given by:

$$f(x) = x^2 = \sum_{n=0}^{\infty} A_n\cos\left(\frac{n\pi x}{L}\right) = A_0 + \sum_{n=0}^{\infty} A_n\cos\left(\frac{n\pi x}{L}\right)$$

$$= \frac{L^2}{3} + \sum_{n=0}^{\infty}\frac{4L^2(-1)^n}{n^2\pi^2}\cos\left(\frac{n\pi x}{L}\right)$$

Note that we will often strip out the $n = 0$ from the series as we have done here because it will almost always be different from the other coefficients and it allows us to actually plug the coefficients into the series.

Now, just as we did in the previous section let's ask what we need to do in order to find the Fourier cosine series of a function that is not even. As with Fourier sine series when we make this change we'll need to move onto the interval $0 \le x \le L$ now instead of $-L \le x \le L$ and again we'll assume that the series will converge to $f(x)$ at this point and leave the discussion of the convergence of this series to a later section.

We could go through the work to find the coefficients here twice as we did with Fourier sine series, however there's no real reason to. So, while we could redo all the work above to get formulas for the coefficients let's instead go straight to the second method of finding the coefficients.

In this case, before we actually proceed with this we'll need to define the even extension of a function, $f(x)$ on $-L \le x \le L$. So, given a function $f(x)$ we'll define the even extension of the function as,

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le L \\ f(-x) & \text{if } -L \le x \le 0 \end{cases}$$

Showing that this is an even function is simple enough.

$$g(-x) = f(-(-x)) = f(x) = g(x) \qquad \text{if } 0 < x < L$$

And we can see that $g(x) = f(x)$ on $0 \le x \le L$ and if $f(x)$ is already an even function we get $g(x) = f(x)$ on $-L \le x \le L$.

Let us take a look at some functions and sketch the even extensions for the functions.

Example 2: Sketch the even extension of each of the given functions.
$f(x) = L - x$ on $0 \le x \le L$.
$f(x) = x^3$ on $0 \le x \le L$

$$f(x) = \begin{cases} \dfrac{L}{2} & \text{if } 0 \le x \le \dfrac{L}{2} \\ x - \dfrac{L}{2} & \text{if } \dfrac{L}{2} \le x \le L \end{cases}$$

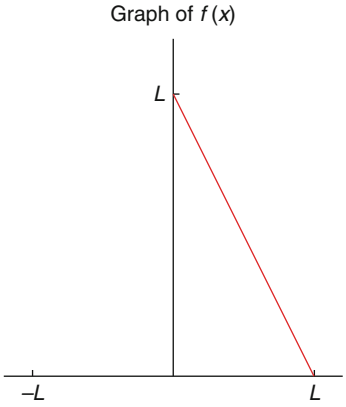
Solution:
(a) $f(x) = L - x$ on $0 \le x \le L$.

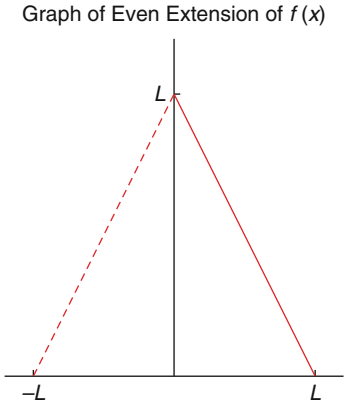
Here is the even extension of this function.

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le L \\ f(-x) & \text{if } -L \le x \le 0 \end{cases}$$

$$= \begin{cases} L - x & \text{if } 0 \le x \le L \\ L + x & \text{if } -L \le x \le 0 \end{cases}$$

Here is the graph of both the original function and its even extension. Note that we have put the "extension" in with a dashed line to make it clear from the portion of the function that is being added to allow us to get the even extension





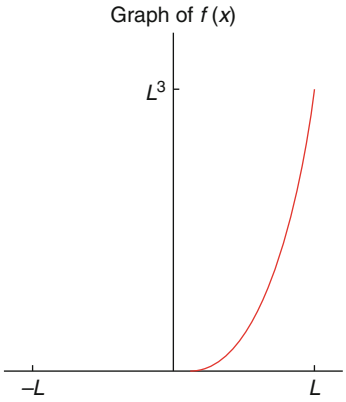
(b) $f(x) = x^3$ on $0 \le x \le L$

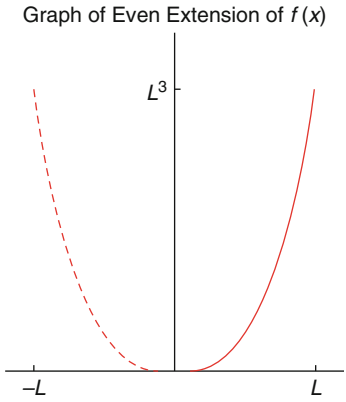
The even extension of this function is

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le L \\ f(-x) & \text{if } -L \le x \le 0 \end{cases}$$

$$= \begin{cases} x^3 & \text{if } 0 \le x \le L \\ -x^3 & \text{if } -L \le x \le 0 \end{cases}$$

The sketch of the function and the even extension is;





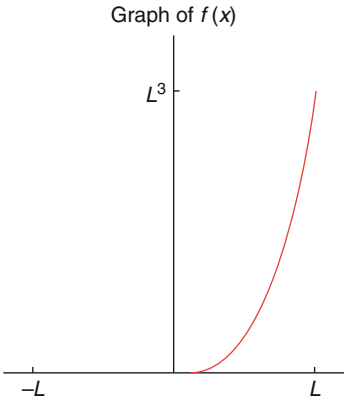
(3)
$$f(x) = \begin{cases} \dfrac{L}{2} & \text{if } 0 \le x \le \dfrac{L}{2} \\ x - \dfrac{L}{2} & \text{if } \dfrac{L}{2} \le x \le L \end{cases}$$

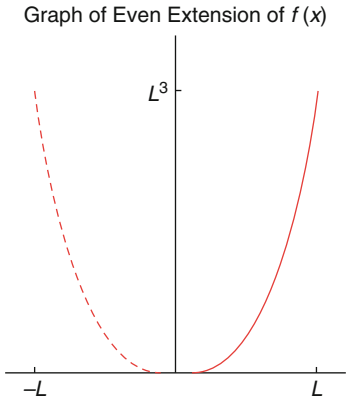
Here is the even extension of this function;

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le L \\ f(-x) & \text{if } -L \le x \le 0 \end{cases}$$

$$= \begin{cases} x - \dfrac{L}{2} & \text{if } \dfrac{L}{2} \le x \le 0 \\ \dfrac{L}{2} & \text{if } 0 \le x \le \dfrac{L}{2} \\ \dfrac{L}{2} & \text{if } -\dfrac{L}{2} \le x \le 0 \\ -x - \dfrac{L}{2} & \text{if } -L \le x \le -\dfrac{L}{2} \end{cases}$$

The sketch of the function and the even extension is;





Now let us think about how we can use the even extension of a function to find the Fourier Cosine Series of any function $f(x)$ on $0 \le x \le L$.

So, given a function $f(x)$ we will let $g(x)$ be the even extension as defined above. Now, $g(x)$ is an even function on $-L \le x \le L$ and so we can write down its Fourier cosine series. This;

$$g(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) \qquad A_n = \begin{cases} \dfrac{1}{L}\displaystyle\int_{-L}^{L} f(x)dx & n = 0 \\ \dfrac{2}{L}\displaystyle\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx & n \neq 0 \end{cases}$$

Let us take a look at a couple of examples.

Example 3: Find the Fourier cosine series for $f(x) = L - x$ on $0 \le x \le L$.

Solution: All we need to do is compute the coefficients, so here is the work for that:

$$A_0 = \frac{1}{L}\int_0^L f(x)dx = \frac{1}{L}\int_0^L (L-x)dx = \frac{L}{2}$$

$$\begin{aligned} A_n &= \frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{2}{L}\int_0^L (L-x)\cos\left(\frac{n\pi x}{L}\right)dx \\ &= \tfrac{2}{L}\left(\frac{L}{n^3\pi^3}\right)\left(n\pi(L-x)\,\sin\left(\frac{n\pi x}{L}\right) - L\cos\left(\frac{n\pi x}{L}\right)\right)\Big|_0^L \\ &= \frac{2}{L}\left(\frac{L}{n^2\pi^2}\right)(-L\cos(n\pi) + L) = \frac{2L}{n^2\pi^2}\left(1 + (-1)^{n+1}\right) \text{ for } n = 1, 2, 3, \ldots \end{aligned}$$

The Fourier cosine series is then

$$f(x) = L - x = \sum_{n=0}^{\infty} \frac{2L}{n^2\pi^2}\left(1 + (-1)^{n+1}\right)\cos\left(\frac{n\pi x}{L}\right)$$

Note that as we did with the first example in this section we stripped out the A_0 term before we plugged in the coefficients.

Next, let us find the Fourier cosine series of an odd function. Note that this is doable because we are really finding the Fourier cosine series of the even extension of the function.

Example 4: Find the Fourier cosine series for $f(x) = x^3$ on $0 \le x \le L$.

Solution: The integral for A_0 is simple enough but the integral for the rest will be fairly messy as it will require three integration by parts. We will leave most of the details of the actual integration to you to verify. Here is the work

$$A_0 = \frac{1}{L}\int_0^L f(x)dx = \frac{1}{L}\int_0^L x^3 dx = \frac{L^3}{4}$$

$$\begin{aligned} A_n &= \frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{2}{L}\int_0^L x^3\cos\left(\frac{n\pi x}{L}\right)dx \\ &= \tfrac{2}{L}\left(\frac{L}{n^4\pi^4}\right)\left(n\pi\left(n^2\pi^2x^2 - 6L^2\right)\sin\left(\frac{n\pi x}{L}\right) - \left(3Ln^2\pi^2x^2 - 6L^3\right)\cos\left(\frac{n\pi x}{L}\right)\right)\Bigg|_0^L \\ &= \frac{2}{L}\left(\frac{L}{n^4\pi^4}\right)\left(n\pi L\left(n^2\pi^2L^2 - 6L^2\right)\sin(n\pi) + \left(3L^3n^2\pi^2 - 6L^3\right)\cos(n\pi) + 6L^3\right) \\ &= \frac{2}{L}\left(\frac{L}{n^4\pi^4}\right)\left(2 + \left(n^2\pi^2 - 2\right)(-1)^n\right) = \frac{6L^3}{n^4\pi^4}\left(2 + \left(n^2\pi^2 - 2\right)(-1)^n\right) \text{ for } n = 1,2,3,\ldots \end{aligned}$$

The Fourier cosine series for this function is then,

$$f(x) = x^3 = \sum_{n=0}^{\infty} \frac{6L^3}{n^4\pi^4}\left(2 + \left(n^2\pi^2 - 2\right)(-1)^{n+1}\right)\cos\left(\frac{n\pi x}{L}\right)$$

Finally, let us take a quick look at a piecewise function.

Example 5: Find the Fourier cosine series of $f(x) = \begin{cases} \dfrac{L}{2} & \text{if } 0 \le x \le \dfrac{L}{2} \\ x - \dfrac{L}{2} & \text{if } \dfrac{L}{2} \le x \le L \end{cases}$ on $0 \le x \le L$.

Solution: We will need to split up the integrals for each of the coefficients here. Here are the coefficients.

$$A_0 = \frac{1}{L}\int_0^L f(x)dx = \frac{1}{4}\left[\int_0^{\frac{L}{2}} f(x)dx + \int_{\frac{L}{2}}^0 f(x)dx\right]$$

$$= \frac{1}{L}\left[\int_0^{\frac{L}{2}} \frac{L}{2}dx + \int_0^{\frac{L}{2}} \left(x - \frac{L}{2}\right)dx\right] = \frac{1}{L}\left[\frac{L^2}{4} + \frac{L^2}{8}\right] = \frac{3L}{8}$$

$$A_n = \frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx$$

$$= \frac{2}{L}\left[\int_0^{\frac{L}{2}} f(x)\cos\left(\frac{n\pi x}{L}\right)dx + \int_{\frac{L}{2}}^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx\right]$$

$$= \frac{2}{L}\left[\int_0^{\frac{L}{2}} \frac{L}{2}\cos\left(\frac{n\pi x}{L}\right)dx + \int_{\frac{L}{2}}^L \left(x - \frac{L}{2}\right)\cos\left(\frac{n\pi x}{L}\right)dx\right]$$

To make calculation a little easier let us do each of these separately.

$$\int_0^{\frac{L}{2}} \frac{L}{2}\cos\left(\frac{n\pi x}{L}\right)dx = \frac{L}{2}\left(\frac{L}{n\pi}\right)\sin\left(\frac{n\pi x}{L}\right)\Bigg|_0^{\frac{L}{2}} = \frac{L}{2}\left(\frac{L}{n\pi}\right)\sin\left(\frac{n\pi}{2}\right) = \frac{L^2}{2n\pi}\sin\left(\frac{n\pi}{2}\right)$$

$$\int_{\frac{L}{2}}^0 \left(x - \tfrac{L}{2}\right)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{L}{n\pi}\left(\frac{L}{n\pi}\cos\left(\frac{n\pi x}{L}\right) + \left(x + \frac{n\pi x}{L}\right)\right)\Bigg|_{\frac{L}{2}}^L$$

$$= \frac{L}{n\pi}\left(\frac{L}{n\pi}\cos(n\pi) - \frac{L}{2}\sin(n\pi) - \frac{L}{n\pi}\cos\left(\frac{n\pi}{2}\right)\right)$$

$$= \frac{L^2}{n^2\pi}\left((-1)^n - \cos\left(\frac{n\pi}{2}\right)\right)$$

Putting these together gives

$$A_n = \frac{2}{L}\left[\frac{L^2}{2n\pi}\sin\left(\frac{n\pi}{2}\right) + \frac{L^2}{n^2\pi^2}\left((-1)^n - \cos\left(\frac{n\pi}{2}\right)\right)\right]$$

So, after all that work the Fourier cosine series is then;

$$f(x) = \frac{3L}{8} + \sum_{n=1}^{\infty} \frac{2L}{n\pi}\left[(-1)^n - \cos\left(\frac{n\pi}{2}\right) + \frac{n\pi}{2}\sin\left(\frac{n\pi}{2}\right)\right]\cos\left(\frac{n\pi x}{L}\right)$$

Note that much as we saw with the Fourier Sine Series many of the coefficients will quite messy to deal with.

E.8 Fourier Series

In the previous two sections we have looked at the Fourier Sine and Fourier Cosine Series. It is now time to look at a Fourier Series. With a Fourier series we are going to try to write a series representation for $f(x)$ on $-L \le x \le L$ in the form;

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=0}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

So a Fourier series is, in some way, a combination of the Fourier sine and Fourier cosine series. Also, like the Fourier sine/cosine series we'll not worry about whether or not the series will actually converge to $f(x)$ or not at this point. For now we'll just assume that it will converge and we'll discuss the convergence of the Fourier series in a later section.

Determining formulas for the coefficients, A_n and B_n, will be done in exactly the same manner as we did in the previous two sections. We will take advantage of the fact that $\left\{\cos\left(\frac{n\pi x}{L}\right)\right\}_{n=0}^{\infty}$ and $\left\{\sin\left(\frac{n\pi x}{L}\right)\right\}_{n=1}$ are mutually orthogonal on $-L \le x \le L$ as we proved earlier. We'll also need the following formulas that we derived when we proved the two sets were mutually orthogonal.

$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = \begin{cases} 2L & \text{if } n = m = 0 \\ L & \text{if } n = m \ne 0 \\ 0 & \text{if } n \ne m \end{cases}$$

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx = \begin{cases} L & \text{if } n = m \\ 0 & \text{if } n \ne m \end{cases}$$

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 0$$

So, let us start off by multiplying both sides of the series above by $\cos\left(\frac{m\pi x}{L}\right)$ and integrating from $-L$ to L. Doing this gives:

$$\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right)dx = \int_{-L}^{L}\sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx + \int_{-L}^{L}\sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx$$

Now, just as we have been able to do in the last two sections we can interchange the integral and the summation. Doing this gives:

$$\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right)dx = \sum_{n=0}^{\infty} A_n \int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx + \sum_{n=1}^{\infty} B_n \int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx$$

We can now take advantage of the fact that the sines and cosines are mutually orthogonal. The integral in the second series will always be zero and in the first series the integral will be zero if and so this reduces to;

$$\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right)dx = \begin{cases} A_m(2L) & \text{if } n = m = 0 \\ A_m(L) & \text{if } n = m \neq 0 \end{cases}$$

Solving for A_m gives;

$$A_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx$$

$$A_m = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right)dx \quad m = 1, 2, 3, \ldots$$

Now, do it all over again only this time multiply both sides by $\sin\left(\frac{m\pi x}{L}\right)$, integrate both sides from $-L$ to L and interchange the integral and summation to get,

$$\int_{-L}^{L} f(x)\sin\left(\frac{m\pi x}{L}\right)dx = \sum_{n=0}^{\infty} A_n \int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx + \sum_{n=1}^{\infty} B_n \int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx$$

In this case the integral in the first series will always be zero and the second will be zero if $n \neq m$ and so we get,

$$\int_{-L}^{L} f(x) \sin\left(\frac{m\pi x}{L}\right) dx = B_m(L)$$

Finally solving for B_m gives,

$$B_m = \frac{1}{L}\int_{-L}^{L} f(x) \sin\left(\frac{m\pi x}{L}\right) dx \quad m = 1, 2, 3, \ldots$$

In the previous two sections we also took advantage of the fact that the integrand was even to give a second form of the coefficients in terms of an integral from 0 to L. However, in this case we don't know anything about whether $f(x)$ will be even, odd, or more likely neither even nor odd. Therefore, this is the only form of the coefficients for the Fourier series.

Before we start with the examples let's remind ourselves of a couple of formulas that we'll make a heavy use of here in this section, as we've done in the previous two sections as well. Provided n in an integer then,

$$\cos(n\pi) = (-1)^n \quad \sin(n\pi)$$

In all of the work that we'll be doing here n will be an integer and so we'll use these without comment in the problems, so be prepared for them.

Also don't forget that sine is an odd function, *i.e.* $\sin(-x) = -\sin(x)$ and that cosine is an even function, *i.e.* $\cos(-x) = \cos(x)$. We will also be making a heavy use of these ideas without comment in many of the integral evaluations, so be ready for these as well.

Summary of Fourier Series:

Let $f(x)$ satisfy the following conditions:
$f(x)$ is defined in the interval $c < x < c + 2L$.
$f(x)$ and $\dot{f}(x)$ are sectionally continuous in $c < x < c + 2L$.
$f(x + 2x) = f(x)$, i.e. $f(x)$ is periodic with period $2L$.
Then at every period of continuity, we have

$$f(x) = \frac{A_0}{2} + \sum_{n=1}^{\infty}\left[A_n \cos\left(\frac{n\pi x}{L}\right) + B_n \sin\left(\frac{n\pi x}{L}\right)\right] \tag{E.3}$$

(continued)

where

$$\left.\begin{aligned} A_n &= \frac{1}{L}\int_c^{c+2L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx \\ B_n &= \frac{1}{L}\int_c^{c+2L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx \end{aligned}\right\} \tag{E.4}$$

At a point of discontinuity, the left side of Eq. E.4 is replaced by $\frac{1}{2}\{f(x+0)+f(x-0)\}$, i.e. the mean value at the discontinuity.

The series Eq. E.3 with coefficients Eq. E.5 is called the Fourier Series of $f(x)$. For many problems, $c = 0$ or $c = -L$. In case $L = \pi$,

Finite Fourier Sine Transform of $f(x), 0 < x < L$, is defined. $f(x)$ has period 2π and Eqs. E.3 and E.4 are simplified.

The above conditions are often called ***Dirichlet Conditions*** and are sufficient (but not necessary conditions for convergence of Fourier Series.

Now let's take a look at an example

Example 1: Find that the Fourier series for $f(x) = L - x$ on $-L \le x \le L$.

Solution: So let us go ahead and just run through formulas for the coefficients.

$$A_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx = \frac{1}{2L}\int_{-L}^{L}(L-x)dx = L$$

$$\begin{aligned} A_n &= \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\int_{-L}^{L}(L-x)\cos\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{1}{L}\left(\frac{L}{n^2\pi^2}\right)\left(n\pi(L-x)\sin\left(\frac{n\pi x}{L}\right) - L\cos\left(\frac{n\pi x}{L}\right)\right)\Bigg|_{-L}^{L} \\ &= \frac{1}{L}\left(\frac{L}{n^2\pi^2}\right)(-2n\pi L\sin(-n\pi)) = 0 \quad n = 1,2,3,\ldots \end{aligned}$$

$$\begin{aligned} B_n &= \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\int_{-L}^{L}(L-x)\sin\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{1}{L}\left(\frac{L}{n^2\pi^2}\right)\left(L\sin\left(\frac{n\pi x}{L}\right) - n\pi(x-L)\cos\left(\frac{n\pi x}{L}\right)\right)\Bigg|_{-L}^{L} \\ &= \frac{1}{L}\left(\frac{L^2}{n^2\pi^2}\right)(2n\pi\cos(n\pi) - 2\sin(n\pi)) = \frac{2L(-1)^n}{n\pi} \quad n = 1,2,3,\ldots \end{aligned}$$

Note that in this case we had $A_0 \neq 0$ and $A_n = 0$ for. $n = 1, 2, 3, \ldots$ This will happen on occasion.

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

$$= A_0 + \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) = L + \sum_{n=1}^{\infty} \frac{2L(-1)^n}{n\pi} \sin\left(\frac{n\pi x}{L}\right)$$

As we saw in the previous example sometime we will get $A_0 \neq 0$ and $A_n = 0$ for. $n = 1, 2, 3, \ldots$. Whether or not this will happen depend upon the function $f(x)$ and often will not happen.

Let us take a look at another example

Example 2: Find the Fourier series for $f(x) = \begin{cases} L & \text{if} -L \leq x \leq \\ 2x & \text{if} 0 \leq x \leq L \end{cases}$ on $-L \leq x \leq L$..

Solution: Because of the piecewise nature of the function the work for the coefficients is going to be a little unpleasant but let us get on with the calculation.

$$A_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx = \frac{1}{2L}\left[\int_{-L}^{0} f(x)dx + \int_{0}^{L} f(x)dx\right]$$

$$= \frac{1}{2L}\left[\int_{-L}^{0} Ldx + \int_{0}^{L} 2xdx\right] = \frac{1}{2L}\left[L^2 + L^2\right] = L$$

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{1}{2L}\left[\int_{-L}^{0} f(x)\cos\left(\frac{n\pi x}{L}\right)dx + \int_{0}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx\right]$$

$$= \frac{1}{2L}\left[\int_{-L}^{0} L\cos\left(\frac{n\pi x}{L}\right)dx + \int_{0}^{L} 2x\cos\left(\frac{n\pi x}{L}\right)dx\right]$$

At this point it will probably be easier to do each of these individually.

$$\int_{-L}^{0} L\cos\left(\frac{n\pi x}{L}\right)dx = \left(\frac{L^2}{n\pi}\sin\left(\frac{n\pi x}{L}\right)\right)\Big|_{-L}^{0} = \frac{L^2}{n\pi}\sin(n\pi) = 0$$

$$\int_0^L 2x\cos\left(\frac{n\pi x}{L}\right)dx = \left(\frac{2L}{n^2\pi^2}\right)\left(L\cos\left(\frac{n\pi x}{L}\right)+n\pi\sin\left(\frac{n\pi x}{L}\right)\right)\Big|_0^L$$

$$= \left(\frac{2L}{n^2\pi^2}\right)(L\cos(n\pi x)+n\pi\sin(n\pi x)-L\cos(0))$$

$$= \left(\frac{2L}{n^2\pi^2}\right)((-1)^n-1)$$

So if we put all of this together we have

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\left[0+\left(\frac{2L^2}{n^2\pi^2}\right)((-1)^n-1)\right]$$

$$= \frac{2L}{n^2\pi^2}((-1)^n-1) \quad n=1,2,3,\ldots$$

So, we have gotten the coefficients for the cosines taken care of and now we need to take care of the coefficients for the sines.

So, we have gotten the coefficients for the cosines taken care of and now we need to take care of the coefficients for the sines.

$$B_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\left[\int_{-L}^{0} f(x)\sin\left(\frac{n\pi x}{L}\right)dx\right]$$

$$= \frac{1}{L}\left[\int_{-L}^{0} L\sin\left(\frac{n\pi x}{L}\right)dx + \int_0^L 2x\sin\left(\frac{n\pi x}{L}\right)dx\right]$$

As with the coefficients for the cosines will probably be easier to do each of these individually.

$$\int_{-L}^{0} L\sin\left(\frac{n\pi x}{L}\right)dx = \left(-\frac{L^2}{n\pi}\cos\left(\frac{n\pi x}{L}\right)\right)\Big|_{-L}^{0} = \frac{L^2}{n\pi}(-1+\cos(n\pi)) = \frac{L^2}{n\pi}((-1)^n-1)$$

$$\int_0^L 2x\sin\left(\frac{2L}{n^2\pi^2}\right)dx = \left(\frac{2L}{n^2\pi^2}\right)\left(L\sin\left(\frac{n\pi x}{L}\right)-n\pi\cos\left(\frac{n\pi x}{L}\right)\right)\Big|_0^L$$

$$= \left(\frac{2L}{n^2\pi^2}\right)\left(L\sin(n\pi)-n\pi\cos\left(\frac{n\pi x}{L}\right)\right)$$

$$= \left(\frac{2L}{n^2\pi^2}\right)(-n\pi(-1)^n) = -\frac{2L^2}{n\pi}(-1)^n$$

So, if we put all of this together we have

$$B_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\left[\frac{L^2}{n\pi}((-1)^n - 1) - \frac{2L^2}{n\pi}(-1)^n\right]$$

$$= \frac{1}{L}\left[\frac{L^2}{n\pi}(-1-(-1)^n)\right] = -\frac{L}{n\pi}(1+(-1)^n) \quad n = 1, 2, 3, \ldots$$

So, after all that work the Fourier series is;

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

$$= A_0 + \sum_{n=1}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

$$= L + \sum_{n=1}^{\infty} \frac{2L}{n^2\pi^2}((-1)^n - 1)\cos\left(\frac{n\pi x}{L}\right) - \sum_{n=1}^{\infty} B_n(1+(-1)^n)\sin\left(\frac{n\pi x}{L}\right)$$

As we saw in the previous example there is often quite a bit of work involved in the computing the integrals involved here.

The next coup a nice observation about some Fourier Series and their relation to Fourier Sine/Cosine Series.

Example 3: Find the Fourier series for $f(x) = x$ on $-L \le x \le L$

Solution: Let us start with the integrals for A_n.

$$A_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx = \frac{1}{2L}\int_{-L}^{L} x dx = 0$$

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\int_{-L}^{L} x\cos\left(\frac{n\pi x}{L}\right)dx = 0$$

In both cases note that we are integrating an odd function (x is odd and cosine is even so the product is odd) over the interval $[-L, L]$ and so we know that both of these integrals will be zero.

Next here is the integral for B_n.

$$B_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\int_{-L}^{L} x\sin\left(\frac{n\pi x}{L}\right)dx = \frac{2}{L}\int_{0}^{L} x\sin\left(\frac{n\pi x}{L}\right)dx$$

In this case we are integrating an even function (x and sine are both odd so the product is even) on the interval $[-L, L]$ so we can "simplify" the integral as shown above. The reason for doing this here is not actually to simplify the integral however. It is instead done so that we can note that we did this integral back in

the Fourier Sine Series section and so do not need to redo it in this section. Using the previous result we get;

$$B_n = \frac{(-1)^{n+1}2L}{n\pi} \quad n = 1, 2, 3, \ldots$$

In this case the Fourier Series is

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}2L}{n\pi} \sin\left(\frac{n\pi x}{L}\right)$$

If you go back and take a look at Example 1 in the Fourier Sine Series section, the same example we used to get the integral out of, you will see that example we were finding the Fourier Sine series for $f(x) = x$ on $-L \le x \le L$. The important thing to note here is that the answer that we got in that example is identical to the answer we got here.

If you think about it however, this should not be too surprising. In both cases we were using an odd function on $-L \le x \le L$ and because we know that we had an odd function the coefficients of the cosines in the Fourier series, A_n, will involve integrating and odd function over a symmetric interval, $-L \le x \le L$, and so will be zero. So, in these cases the Fourier sine series of an odd function on $-L \le x \le L$ is really just a special case of a Fourier series.

Note however that when we moved over to doing the Fourier sine series of any function on $-L \le x \le L$ we should no longer expect to get the same results. You can see this by comparing Example 1 above with Example 3 in the Fourier sine series section. In both examples we are finding the series for $f(x) = x - L$ and yet got very different answers.

So, why did we get different answers in this case?. Recall that when we find the Fourier sine series of a function on $0 \le x \le L$ we are really finding the Fourier sine series of the odd extension of the function on $-L \le x \le L$ and then just restricting the result down to $0 \le x \le L$. For a Fourier series we are actually using the whole function on $-L \le x \le L$ instead of its odd extension. We should therefore not expect to get the same results since we are really using different functions (at least on part of the interval) in each case.

So, if the Fourier sine series of an odd function is just a special case of a Fourier series it makes some sense that the Fourier cosine series of an even function should also be a special case of a Fourier series. Let's do a quick example to verify this.

Example 4: Find the Fourier series for $f(x) = x^2$ on $-L \le x \le L$.

Solution: Here are the integrals for the A_n and in this case because both the function and cosine are even we will be integration an even function and so can "simplify" the integral.

$$A_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx = \frac{1}{2L}\int_{-L}^{L} x^2 dx$$

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{1}{L}\int_{-L}^{L} x^2\cos\left(\frac{n\pi x}{L}\right)dx = \frac{2}{L}\int_{0}^{L} x^2\cos\left(\frac{n\pi x}{L}\right)dx$$

As with the previous example both of these integrals were done in Example 1 in the Fourier Cosine Series and so we will not bother redoing them here. The coefficient are;

$$A_0 = \frac{L^2}{3} \quad A_n = \frac{4L^2(-1)^n}{n^2\pi^2} \quad n = 1, 2, 3, \ldots$$

Next there is the integral for B_n

In this case the function is even and sine is odd so the product is odd and we are integrating over $-L \le x \le L$ and so the integral is zero.

The Fourier Series is then;

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) = \frac{L^3}{3} + \sum_{n=1}^{\infty} \frac{4L^2(-1)^n}{n^2\pi^2}\cos\left(\frac{n\pi x}{L}\right)$$

As suggested before we started this example the result here is identical to the result from Example 1 in the Fourier cosine series section and so we can see that the Fourier cosine series of an even function is just a special case a Fourier series.

E.9 Convergence of Fourier Series

Over the last few sections we've spent a fair amount of time to computing Fourier series, but we have avoided discussing the topic of convergence of the series. In other words, will the Fourier series converge to the function on the given interval?

In this section we're going to address this issue as well as a couple of other issues about Fourier series. We'll be giving a fair number of theorems in this section but are not going to be proving any of them. We'll also not be doing a whole lot of in the way of examples in this section.

Before we get into the topic of convergence we need to first define a couple of terms that we'll run into in the rest of the section. First, we say that $f(x)$ has a **jump discontinuity** at $x = a$ if the limit of the function from the left, denoted $f(a^-)$, and the limit of the function from the right, denoted $f(a^+)$, both exist and $f(a^-) \neq f(a^+)$.

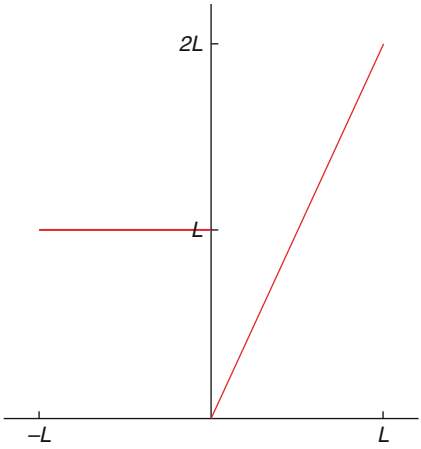
Next, we say that $f(x)$ is **piecewise smooth** if the function can be broken into distinct pieces and on each piece both the function and its derivative, $f'(x)$, are

continuous. A piecewise smooth function may not be continuous everywhere however the only discontinuities that are allowed are a finite number of jump discontinuities.

Let's consider the function,

$$f(x) = \begin{cases} L & \text{if } -L \le x \le 0 \\ 2x & \text{if } 0 \le x \le L \end{cases}$$

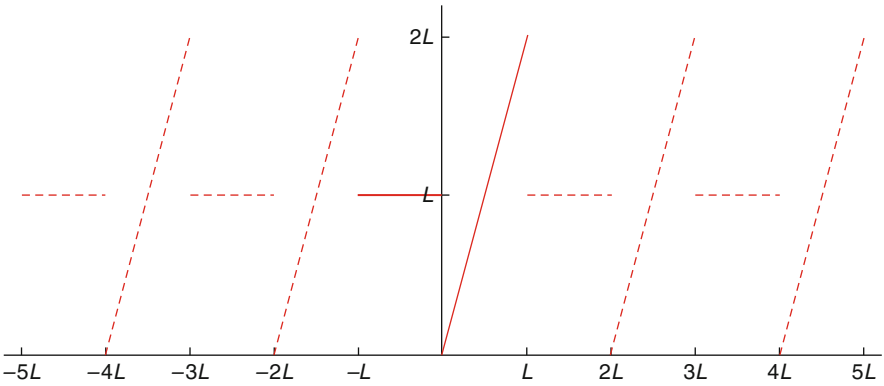
We found the Fourier series for this function in Example 2 of the previous section. Here is sketch of this function on the interval on which it is defined, *i.e.* $-L \le x \le L$.



This function has a jump discontinuity at $x = 0$ because $f(0^-) = L \ne 0 = f(0^+)$ and note that on the intervals $-L \le x \le L$ and $0 \le x \le L$ both the function and its derivative are continuous. This is therefore an example of a piecewise smooth function. Note that the function itself is not continuous at $x = 0$ but because this point of discontinuity is a jump discontinuity the function is still piecewise smooth.

The last term we need to define is that of periodic extension. Given a function, $f(x)$, defined on some interval, we'll be using $-L \le x \le L$ exclusively here, the periodic extension of this function is the new function we get by taking the graph of the function on the given interval and then repeating that graph to the right and left of the graph of the original function on the given interval.

It is probably best to see an example of a periodic extension at this point to help make the words above a little clearer. Here is a sketch of the period extension of the function we looked at above,



The original function is the solid line in the range $-L \le x \le L$. We then got the periodic extension of this by picking this piece up and copying it every interval of length $2L$ to the right and left of the original graph. This is shown with the two sets of dashed lines to either side of the original graph.

Note that the resulting function that we get from defining the periodic extension is in fact a new periodic function that is equal to the original function on $-L \le x \le L$.

With these definitions out of the way we can now proceed to talk a little bit about the convergence of Fourier series. We will start off with the convergence of a Fourier series and once we have that taken care of the convergence of Fourier Sine/Cosine series will follow as a direct consequence. Here then is the theorem giving the convergence of a Fourier series.

Convergence of Fourier Series

Suppose $f(x)$ is a piecewise smooth on the $-L \le x \le L$. The Fourier Series of $f(x)$ will then converge to;

The periodic extension of $f(x)$ if the periodic extension is continuous.

The average of the two one-sided limit, $\frac{1}{2}[f(a^-) + f(a^+)]$, if the periodic extension has a jump discontinuity at $x = a$.

The first thing to note about this is that on the interval $-L \le x \le L$ both the function and the periodic extension are equal and so where the function is continuous on $-L \le x \le L$ the periodic extension will also be continuous and hence at these points the Fourier series will in fact converge to the function. The only points in the interval $-L \le x \le L$ where the Fourier series will not converge to the function is where the function has a jump discontinuity.

Let's again consider Example 2 of the previous section. In that section we found that the Fourier series of,

$$f(x) = \begin{cases} L & \text{if } -L \le x \le 0 \\ 2x & \text{if } 0 \le x \le L \end{cases}$$

on $-L \le x \le L$ to be,

$$f(x) = L + \sum_{n=1}^{\infty} \frac{2L}{n^2\pi^2}\left((-1)^n - 1\right)\cos\left(\frac{n\pi x}{L}\right) - \sum_{n=1}^{\infty} \frac{L}{n\pi}(1 + (-1)^n)\sin\left(\frac{n\pi x}{L}\right)$$

We now know that in the intervals $-L < x < 0$ and $0 < x < L$ the function and hence the periodic extensions are both continuous and so on these two intervals the Fourier series will converge to the periodic extension and hence will converge to the function itself.

At the point $x = 0$ the function has a jump discontinuity and so the periodic extension will also have a jump discontinuity at this point. That means that at $x = 0$ the Fourier series will converge to,

$$\frac{1}{2}\left[f(0^-) + f(0^+)\right] = \frac{1}{2}[L + 0] = \frac{L}{2}$$

At the two endpoints of the interval, $x = -L$ and $x = L$, we can see from the sketch of the periodic extension above that the periodic extension has a jump discontinuity here and so the Fourier series will not converge to the function there but instead the averages of the limits.

So, at $x = -L$ the Fourier series will converge to,

$$\frac{1}{2}\left[f(-L^-) + f(-L^+)\right] = \frac{1}{2}[2L + L] = \frac{3L}{2}$$

and at $x = L$ the Fourier series will converge to,

$$\frac{1}{2}\left[f(L^-) + f(L^+)\right] = \frac{1}{2}[2L + L] = \frac{3L}{2}$$

Now that we have addressed the convergence of a Fourier series we can briefly turn our attention to the convergence of Fourier sine/cosine series. First, as noted in the previous section the Fourier sine series of an odd function on $-L \le x \le L$ and the Fourier cosine series of an even function on are both just special cases of a Fourier series we now know that both of these will have the same convergence as a Fourier series.

Next, if we look at the Fourier sine series of any function, $g(x)$, on $0 \le x \le L$ then we know that this is just the Fourier series of the odd extension of restricted down to the interval. Therefore we know that the Fourier series will converge to the odd extension on $-L \le x \le L$ where it is continuous and the average of the limits where the odd extension has a jump discontinuity. However, on $0 \le x \le L$ we know that

$g(x)$ and the odd extension are equal and so we can again see that the Fourier sine series will have the same convergence as the Fourier series.

Likewise, we can go through a similar argument for the Fourier cosine series using even extensions to see that Fourier cosine series for a function on $0 \le x \le L$ will also have the same convergence as a Fourier series.

The next topic that we want to briefly discuss here is when will a Fourier Series be continuous. From the theorem on the convergence of Fourier series we know that where the function is continuous the Fourier series will converge to the function and hence be continuous at these points. The only place where the Fourier Series may not be continuous is if there is a jump discontinuity on the interval $-L \le x \le L$ and potentially at the endpoints as we saw that the periodic extension may introduce a jump discontinuity there.

So, if we're going to want the Fourier Series to be continuous everywhere we'll need to make sure that the function does not have any discontinuities in $-L \le x \le L$. Also, in order to avoid having the periodic extension introduce a jump discontinuity we'll need to require that $f(-L) = f(L)$. By doing this the two ends of the graph will match up when we form the periodic extension and hence we will avoid a jump discontinuity at the end points.

Here is a summary of these ideas for a Fourier Series.

Suppose $f(x)$ is a piecewise smooth on the $-L \le x \le L$. The Fourier Series of $f(x)$ will be continuous and will converge to $f(x)$ on $-L \le x \le L$ provided $f(x)$ is continuous on $-L \le x \le L$ and $f(-L) = f(L)$.

Now, how can we use this to get similar statements about Fourier sine/cosine series on $0 \le x \le L$. Let's start with a Fourier cosine series. The first thing that we do is form the even extension of $f(x)$ on $-L \le x \le L$. For the purposes of this discussion let's call the even extension $g(x)$. As we saw when we sketched several even extensions in the Fourier cosine series section that in order for the sketch to be the even extension of the function we must have both,

$$g(0^-) = g(0^+) \quad g(-L) = g(L)$$

If one or both of these aren't true then $g(x)$ will not be an even extension of $f(x)$.

So, in forming the even extension we do not introduce any jump discontinuities at $x = 0$ and we get for free that $g(-L) = g(L)$. If we now apply the above theorem to the even extension we see that the Fourier series of the even extension is continuous on $-L \le x \le L$. However, because the even extension and the function itself are the same on $0 \le x \le L$ then the Fourier cosine series of $f(x)$ must also be continuous on $0 \le x \le L$.

Here is a summary of this discussion for the Fourier cosine series.

Suppose $f(x)$ is a piecewise smooth on the. $0 \le x \le L$. The Fourier Series of $f(x)$ will be continuous and will converge to $f(x)$ on $0 \le x \le L$ provided $f(x)$ is continuous on $0 \le x \le L$.

Note that we don't need any requirements on the end points here because they are trivially satisfied when we convert over to the even extension.

For a Fourier sine series we need to be a little more careful. Again, the first thing that we need to do is form the odd extension on $-L \le x \le L$ and let's call it $g(x)$. We know that in order for it to be the odd extension then we know that at all points in $-L \le x \le L$ it must satisfy $g(-x) = -g(x)$ and that is what can lead to problems.

As we saw in the Fourier Sine Series section it is very easy to introduce a jump discontinuity at when we form the odd extension. In fact, the only way to avoid forming a jump discontinuity at this point is to require that $f(0) = 0$.

Next, the requirement that at the endpoints we must have $g(-L) = -g(L)$ will practically guarantee that we'll introduce a jump discontinuity here as well when we form the odd extension. Again, the only way to avoid doing this is to require $0 \le x \le L$.

So, with these two requirements we will get an odd extension that is continuous and so we know that the Fourier Series of the odd extension on $-L \le x \le L$ will be continuous and hence the Fourier Sine Series will be continuous on $f(L) = 0$.

Here is a summary of all this for the Fourier sine series.

Suppose $f(L)$ is a piecewise smooth on the. $0 \le x \le L$. The Fourier Series of $f(x)$ will be continuous and will converge to $f(x)$ on $0 \le x \le L$ provided $f(x)$ is continuous on $0 \le x \le L, f(0) = 0$ and $f(L) = 0$.

The next topic of discussion here is differentiation and integration of Fourier series. In particular we want to know if we can differentiate a Fourier series term by term and have the result be the Fourier series of the derivative of the function. Likewise we want to know if we can integrate a Fourier series term by term and arrive at the Fourier Series of the integral of the function.

Note that we'll not be doing much discussion of the details here. All we're really going to be doing is giving the theorems that govern the ideas here so that we can say we've given them.

Let's start off with the theorem for term by term differentiation of a Fourier Series.

Given a function $f(x)$ if the derivative $\dot{f}(x)$ is a piecewise smooth and the Fourier series of is continuous then the Fourier Series can be differentiated term by term. The result of the differentiation is the Fourier Series of the derivative, $f(x)$.

$f(-L) = f(L)$ then the Fourier series of the function can be differentiated term by term and the results is the Fourier series of the derivative.

One of the main conditions of this theorem is that the Fourier series be continuous and from above we also know the conditions on the function that will give this. So, if we add this into the theorem to get this form of the theorem,

Suppose $f(x)$ is a continuous function, its derivative $\dot{f}(x)$ is a piecewise smooth and $f(-L) = f(L)$ then the Fourier series of the function can be differentiated term by term and the results is the Fourier series of the derivative.

For Fourier Cosine/Sine Series the basic theorem is the same as for Fourier series. All that's required is that the Fourier cosine/sine series is continuous and then you can differentiate term by term. The theorems that we'll give here will merge the conditions for the Fourier cosine/sine series to be continuous into the theorem.

Let's start with the Fourier Cosine Series.

Suppose $f(x)$ is a continuous function and its derivative $\dot{f}(x)$ is a piecewise smooth then the Fourier Cosine Series of the function can be differentiated term by term and the results is the Fourier Sine Series of the derivative.

Next the theorem for Fourier Sine Series

Suppose $f(x)$ is a continuous function, its derivative $\dot{f}(x)$ is a piecewise smooth, $f(0) = 0$ and $f(L) = 0$ then the Fourier Sine Series of the function can be differentiated term by term and the results is the Fourier Cosine Series of the derivative.

The theorem for integration of Fourier series term by term is simple so there it is.

Suppose $f(x)$ is a piecewise smooth then the Fourier Sine Series of the function can be integrated term by term and the result is a convergent infinite series that will converge to the integral of $f(x)$.

Note however that the new series that results from term by term integration may not be the Fourier series for the integral of the function.

E.10 Half Range Fourier Sine and Cosine Series

A half range Fourier sine or cosine series is a series in which only sine terms or only cosine terms are present respectively. When a half range series corresponding to a given function is desired; the function is generally defined in the interval $(0, L)$ [which is half of the interval $(-L, L)$, thus accounting for the name *half range*] and then the function is specified as odd or even, so that it is clearly defined in the other half of the interval, namely $(-L, 0)$. In such case, we have

$$A_n = 0, \text{ then } B_n = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right)dx \text{ for half range sine series} \tag{E.5}$$

$$B_n = 0, \text{ then } A_n = \frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx \text{ for half range sine series}$$

E.11 Complex Form of Fourier Series

In complex notation, the Fourier series Eq. E.3 and coefficients Eq. E.4 of Sect. 1.7 in Appendix E can be written as

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{\frac{in\pi x}{L}} = \sum_{n=-\infty}^{\infty} C_n\left[\cos\left(\frac{n\pi x}{L}\right) + i\sin\left(\frac{n\pi x}{L}\right)\right] \tag{E.6}$$

where $c = -L$, and

$$C_n = \frac{1}{2L}\int_{-L}^{L} f(x)e^{\frac{in\pi x}{L}}dx = \frac{1}{2L}\int_{-L}^{L} f(x)\left[\cos\left(\frac{in\pi x}{L}\right) - i\sin\left(\frac{in\pi x}{L}\right)\right]dx \tag{E.7}$$

E.12 Parseval's Identify for Fourier Series

Parseval's identity states that

$$\frac{1}{L}\int_{-L}^{L}\{f(x)\}^2 dx = \frac{A_0^2}{2} + \sum_{n=1}^{\infty}\left(A_0^2 + B_n^2\right) \tag{E.8}$$

where A_n and B_n are given by Eq. E.8.

An important consequences is that

$$\left.\begin{aligned}\lim_{x\to\infty}\int_{-L}^{L} f(x)\sin\left(\frac{in\pi x}{L}\right)=0\\ \lim_{x\to\infty}\int_{-L}^{L} f(x)\cos\left(\frac{in\pi x}{L}\right)=0\end{aligned}\right\} \tag{E.9}$$

This is called *Riemann's theorem*.

E.13 Finite Fourier Transform

Let $f(x)$ satisfy the following conditions:

1. $f(x)$ is defined in the interval $c < x < c + 2L$.
2. $f(x)$ and $f'(x)$ are sectionally continuous in $c < x < c + 2L$.
3. $f(x + 2x) = f(x)$, i.e. $f(x)$ is periodic with period $2L$.

Then at every period of continuity, we have

$$f(x)=\frac{A_0}{2}+\sum_{n=1}^{\infty}\left[A_n\cos\left(\frac{n\pi x}{L}\right)+B_n\sin\left(\frac{n\pi x}{L}\right)\right] \tag{E.10}$$

where

$$\left.\begin{aligned}A_L&=\frac{1}{L}\int_c^{c+2L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx\\ B_n&=\frac{1}{L}\int_c^{c+2L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx\end{aligned}\right\} \tag{E.11}$$

At a point of discontinuity, the left side of Eq. E.11 is replaced by $\frac{1}{2}\{f(x+0)+f(x-0)\}$, i.e., the mean value at the discontinuity.

The series Eq. E.10 with coefficients Eq. E.11 is called the Fourier Series of $f(x)$. For many problems, $c = 0$ or $c = -L$. In case $L = \pi$,

Finite Fourier Sine Transform of $f(x)$, $0 < x < L$, is defined. $f(x)$ has period 2π and Eqs. E.3 and E.4 are simplified.

The above conditions are often called ***Dirichlet Conditions*** and are sufficient (but not necessary conditions for convergence of Fourier Series.

The *Finite Fourier Sine Transform* of $f(x)$, $0 < x < L$, is defined as.

$$\mathbb{F}_{\text{sine}}(n) = \mathbb{F}_{\text{sine}}\{f(x)\} = \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx \tag{E.12}$$

The *Inverse Finite Fourier Sine Transform* of $\mathbb{F}_{\text{sine}}(n)$ for integer n.

$$f(x) = \frac{2}{L} \sum_{n=1}^{\infty} \mathbb{F}_{\text{sine}}(n) \sin \frac{n\pi x}{L} \tag{E.13}$$

The *Finite Fourier Cosine Transform* of $f(x)$, $0 < x < L$, is defined as.

$$\mathbb{F}_{\text{cosine}}(n) = \mathbb{F}_{\text{cosine}}\{f(x)\} = \int_0^L f(x) \cos\left(\frac{n\pi x}{L}\right) dx \tag{E.14}$$

The *Inverse Finite Fourier Cosine Transform* of $\mathbb{F}_{\text{cosine}}(n)$ for integer n.

$$f(x) = \frac{1}{L} \mathbb{F}_{\text{cosine}}(0) + \frac{2}{L} \sum_{n=1}^{\infty} \mathbb{F}_{\text{cosine}}(n) \cos \frac{n\pi x}{L} \tag{E.15}$$

Example 1: Establish both, (a) Eq. E.13 and (b) Eq. E.15 in above

Solution: (a): If $f(x)$ is an odd function in $(-L, L)$, then

$$f(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) \tag{E.16}$$

where

$$B_n = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx \tag{E.17}$$

Thus if we write

$$\int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx = f_{\text{sine}}(n)$$

Then $B_n = \frac{2}{L} f_{\text{sine}}(n)$ and Eq. (E.16) can be written as required,

$$f(x) = \frac{2}{L} \sum_{n=1}^{\infty} f_{\text{sine}}(n) \sin\left(\frac{n\pi x}{L}\right) \tag{E.18}$$

We can also write $f(x) = \mathbb{F}_{\text{sine}}{}^{-1}\{f_{\text{sine}}(n)\}$.

(b): If $f(x)$ is an odd function in $(-L, L)$, then

$$f(x) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos\left(\frac{n\pi x}{L}\right) \tag{E.19}$$

where

$$A_n = \frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx \tag{E.20}$$

Thus if we write

$$\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx = f_{\text{cosine}}(n)$$

then $A_0 = \dfrac{2}{L} f_{\text{cosine}}(0)$ and Eq. (E.19) can be written, as required,

$$f(x) = \frac{1}{L} f_{\text{cosine}}(0) + \frac{2}{L}\sum_{n=1}^{\infty} f_{\text{cosine}}(n)\cos\left(\frac{n\pi x}{L}\right) \tag{E.21}$$

We can also write $f(x) = \mathbb{F}^{-1}_{\text{cosine}}\{f_{\text{cosine}}(n)\}$.

Example 2: Find the (a) finite Fourier sine transform and (b) finite Fourier cosine transform of the function $f(x) = 2x$ for interval $0 < x < 4$.

Solution: (a): Since $L = 4$, we have

$$f_{\text{sine}}(n) = \int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right)dx = \int_0^4 2x\sin\left(\frac{n\pi x}{4}\right)dx$$

$$= \left\{(2x)\left(\frac{-\cos(n\pi x/4)}{n\pi/4}\right) - (2)\left(\frac{-\sin(n\pi x/4)}{n^2\pi^2/16}\right)\right\}\Bigg|_0^4 = -\frac{32}{n\pi}\cos n\pi$$

(b): If $n > 0$, then we have

$$f_{\text{cosine}}(n) = \int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \int_0^4 2x\cos\left(\frac{n\pi x}{4}\right)dx$$

$$= \left\{(2x)\left(\frac{-\sin(n\pi x/4)}{n\pi/4}\right) - (2)\left(\frac{-\cos(n\pi x/4)}{n^2\pi^2/16}\right)\right\}\Bigg|_0^4 = 32\left(\frac{\cos(n\pi) - 1}{n^2\pi^2}\right)$$

If $n = 0$, the we have

$$f_{\text{cosine}}(n) = f_{\text{cosine}}(0) = \int_0^4 2x dx = 16$$

Example 3: Find $f(x)$ if:

(a): $\mathbb{F}_{\text{sine}}\{f(x)\} = 16(-1)^{n-1}/n^3$, $n = 1, 2, 3, \ldots$, where $0 < x < 8$;

(b): $\mathbb{F}_{\text{cosine}}\{f(x)\} \sin(n\pi/2/2n)$, $n = 1, 2, 3, \ldots$ and $\pi/2$ if $n = 0$, where $0 < x < 2\pi$

Solution: (a): From Eq. (E.18) of Example 1 (a) above with $L = 8$, we have

$$f(x) = \mathbb{F}^{-1}_{\text{sine}}\left\{\frac{16(-1)^{n-1}}{n^3}\right\}$$

$$= \frac{2}{8}\sum_{n=1}^{\infty}\frac{16(-1)^{n-1}}{n^3}\sin\frac{n\pi x}{8} = 4\sum_{n=1}^{\infty}\frac{16(-1)^{n-1}}{n^3}\sin\frac{n\pi x}{8}$$

(b): From Eq. (E.21) of Example 1 (b) with $L = 2\pi$, we have

$$f(x) = \mathbb{F}^{-1}_{\text{sine}}\left\{\frac{\sin(n\pi/2)}{2n}\right\}$$

$$= \frac{1}{\pi}\cdot + \frac{2}{2\pi}\sum_{n=1}^{\infty}\frac{\sin(n\pi/2)}{2n} = \frac{1}{4} + \frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{\sin(n\pi/2)}{n}$$

E.14 The Fourier Integral

Let $f(x)$ satisfy the following conditions:

1. $f(x)$ satisfies the ***Dirichlet Condition*** (See Sect. 1.7 of Appendix E) in every finite interval $-L \le x \le L$.
2. $\int_{-L}^{L} |f(x)|dx$ converges, i.e. $f(x)$ is absolutely integrable in $-\infty < x < \infty$.

Then ***Fourier Integral Theorem*** states that

$$f(x) = \int_0^{\infty} \{A(\lambda)\cos\lambda x + B(\lambda)\sin\lambda x\}d\lambda \tag{E.22}$$

where

$$\left.\begin{aligned} A(\lambda) &= \frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\cos\lambda x dx \\ B(\lambda) &= \frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\sin\lambda x dx \end{aligned}\right\} \tag{E.23}$$

This can be written equivalently as

$$f(x) = \frac{2}{\pi}\int_{\lambda=-}^{\infty}\int_{u=-\infty}^{\infty} f(u)\cos\lambda(x-u)du d\lambda \tag{E.24}$$

The result Eq. E.22 holds if x is a point of continuity of $f(x)$. If is a point of discontinuity, we must replace $f(x)$ by $\frac{1}{2}\{f(x+0)+f(x-0)\}$ as in the case of Fourier Series. As for Fourier Series, the above conditions are sufficient but not necessary. The similarity of Eqs. E.22 and E.23 with corresponding results of Eqs. E.3 and E.4 for Fourier Series is apparent. The right side of Eq. E.22 is sometimes called the *Fourier Integral Expansion* of $f(x)$, or *briefly Fourier Integral*. To prove Fourier Integral Theorem we look at the following example.

Example 1: Show a logical demonstration of Fourier's Integral Theorem by use of a limiting form of Fourier Series.

Solution: We write Fourier series of function $f(x)$ to be as follows:

$$f(x) = \frac{A_0}{2} + \sum_{n=1}^{\infty}\left[A_n\cos\left(\frac{n\pi x}{L}\right) + B_n\sin\left(\frac{n\pi x}{L}\right)\right] \tag{E.25}$$

where $A_n = \frac{1}{L}\int_{-L}^{L} f(u)\cos\left(\frac{n\pi x}{L}\right)du$ and $B_n = \frac{1}{L}\int_{-L}^{L} f(u)\sin\left(\frac{n\pi x}{L}\right)du$

Then by substituting of these coefficients into Eq. (E.25) and using result of Example 1, in Sect. 1.12 in above, we find

$$f(x) = \frac{1}{2L}\int_{-L}^{L} f(u)du + \frac{1}{L}\sum_{n=1}^{\infty} f(u)\cos\frac{n\pi}{L}(u-x)du \tag{E.26}$$

If we assume that $\int_{-\infty}^{\infty}|f(u)|du$ converges, the first term on the right of Eq. (E.26) approaches zero as $L\to\infty$, while the remaining part appears to approach

$$\lim_{L\to\infty}\frac{1}{L}\sum_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u)\cos\frac{n\pi}{L}(u-x)du \tag{E.27}$$

This last step is not rigorous and makes the demonstration of this example logical. Calling $\Delta a = \pi/L$, Eq. (E.27) can be written

$$f(x) = \lim_{\Delta a\to 0}\sum_{n=1}^{\infty}\Delta\alpha f(n\Delta\alpha) \tag{E.28}$$

where we have written.

$$f(\alpha) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(u)\cos\alpha(u-x)du \tag{E.29}$$

But the limit Eq. (E.28) is equal to

$$f(x) = \int_0^{\infty} f(\alpha)d\alpha = \frac{1}{\pi}\int_0^{\infty} d\alpha \int_{-\infty}^{\infty} f(u)\cos(u-x)d\alpha \tag{E.30}$$

Which is Fourier's integral formula.

E.15 Complex Form of Fourier Integrals

In complex notation, the Fourier integral Eq. E.22 with coefficients Eq. E.23 can be written as

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{\lambda=-\infty}^{\infty} e^{i\lambda x}d\lambda \int_{u=-\infty}^{\infty} f(u)e^{i\lambda u} \\ &= \frac{1}{2\pi}\int_{\lambda=-\infty}^{\infty}\int_{u=-\infty}^{\infty} f(u)e^{i\lambda(x-u)}dud\lambda \end{aligned} \tag{E.31}$$

To prove that we can again use Eqs. E.3 and E.4 and define cosines and sines in Eq. E.4 via Euler formulas (See Appendix D Sect. 1.6) with exponential function, the series Eq. E.3 attains the form of the following;

$$f(x) = \sum_{n=1}^{\infty} C_n e^{\frac{in\pi x}{L}} \tag{E.32}$$

The coefficient C_n could be obtained of A_n and B_n, but they are comfortably derive directly by multiplying the by $e\dfrac{in\pi x}{L}$ and integrating it from $-L$ to L one obtains

$$C_n = \int_{-L}^{L} f(x)e^{-\frac{in\pi x}{L}}dx \quad n = 0, \pm 1, \pm 2, ... \tag{E.33}$$

Note: We may say that in Eq. E.32, $f(x)$ has been dissolved to sum of *harmonics* (elementary waves) $C_n e^{\frac{in\pi x}{L}}$ with amplitudes C_n corresponding the frequency n.

Derivation will start with seeing the expansion Eq. E.32 changes when $L \to \infty$, we put first the Eq. E.33 into Eq. E.32 substituting for C_n:

$$f(x) = \sum_{n=-\infty}^{\infty} e^{\frac{in\pi x}{L}} \int_{-L}^{L} f(x) e^{-\frac{in\pi x}{L}} dx$$

By denoting $\omega_n = \frac{n\pi}{L}$ and $\Delta\omega_n = \omega_{n+1} - \omega_n = \frac{(n+1)\pi}{L} - \frac{\pi}{L} = \frac{\pi}{L}$, the last equation in above takes the form

$$f(x) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{i\omega_n x} \Delta\omega_n \int_{-L}^{L} f(x) e^{-i\omega_n x} dx$$

It can be shown that when $L \to \infty$ and thus $\Delta\omega_n \to 0$, the limiting form of this equation is

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} d\omega \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \tag{E.34}$$

Here $f(x)$ has been represented as a *Fourier Integral*. It can be proved that for validity of expansion Eq. E.33 it suffices that the function $f(x)$ is piecewise continuous on every finite interval having at most a finite amount of extremum points and that the integral $\int_{-\infty}^{\infty} |f(x)| dx$ converges.

For better to compare to the Fourier Series Eq. E.32 and the coefficient Eq. E.33, we can write Eq. E.34 as;

$$f(x) = \int_{-\infty}^{\infty} C(\omega) f(x) e^{i\omega x} dx$$

where

$$C(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$$

E.16 Fourier Transforms

From Eq. E.31 it follows that if

$$f(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda u} f(u) du \tag{E.35}$$

then

$$f(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\lambda u} f(\lambda) d\lambda \tag{E.36}$$

which gives $f(x)$ on replacing u by x as follows:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\lambda x} f(\lambda) d\lambda \tag{E.37}$$

The function of $f(x)$ is called the *Fourier Transform* of $f(x)$ and id sometimes written as $f(\lambda) = \mathbb{F}\{f(x)\}$. The function $f(x)$ is the *Inverse Fourier Transform* of $f(\lambda)$ and is written $f(x) = \mathbb{F}^{-1}\{f(\lambda)\}$. We also call Eq. E.37 an *Inversion Formula* corresponding to Eq. E.35.

Note that the constants preceding the integral signs can be constants whose product is $1/2\pi$. If they are each taken as $1/\sqrt{2\pi}$ we obtain the so-called *symmetric form*.

E.17 Finite and Infinite Fourier Sine and Cosine Transforms

1. The *Finite Fourier Sine Transform* of $f(x)$, $0 < x < L$, is defined as.

$$\mathbb{F}_{\text{sine}}(n) = \mathbb{F}_{\text{sine}}\{f(x)\} = \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx$$

2. The *Inverse Finite Fourier Sine Transform* of $\mathbb{F}_{\text{sine}}(n)$ for integer n.

$$f(x) = \frac{2}{L}\sum_{n=1}^{\infty} \mathbb{F}_{\text{sine}}(n) \sin\frac{n\pi x}{L}$$

3. The *Finite Fourier Cosine Transform* of $f(x)$, $0 < x < L$, is defined as.

$$\mathbb{F}_{\text{cosine}}(n) = \mathbb{F}_{\text{cosine}}\{f(x)\} = \int_0^L f(x) \cos\left(\frac{n\pi x}{L}\right) dx$$

4. The *Inverse Finite Fourier Cosine Transform* of $\mathbb{F}_{\text{cosine}}(n)$ for integer n.

$$f(x) = \frac{1}{L}\mathbb{F}_{\text{cosine}}(0) + \frac{2}{L}\sum_{n=1}^{\infty}\mathbb{F}_{\text{cosine}}(n)\cos\frac{n\pi x}{L}$$

1. The *Inverse Finite Fourier Cosine Transform* of $\mathrm{F}_{\text{cosine}}(n)$ for integer n.

$$f(x) = \frac{1}{L}\mathbb{F}_{\text{cosine}}\frac{2}{L}\sum_{n=1}^{\infty}\mathbb{F}_{\text{cosine}}(n)\cos\frac{n\pi x}{L}$$

2. The (*infinite*) Fourier Sine Transform of $f(x)$, $0 < x < \infty$, is defined as

$$\mathbb{F}_{\text{sine}}(\lambda)\mathbb{F}_{\text{sine}}\{f(x)\} = \int_0^{\infty} f(x)\sin(\lambda x)dx$$

The function $f(x)$ is then called the *Inverse Fourier Sine Transform* of $\mathbb{F}_{\text{sine}}(\lambda)$ and is given by:

$$f(x) = \frac{2}{\pi}\int_0^{\infty}\mathbb{F}_{\text{sine}}(\lambda)\sin(\lambda x)dx$$

3. The (*infinite*) Fourier Cosine Transform of $f(x)$, $0 < x < \infty$, is defined as

$$\mathbb{F}_{\text{cosine}}(\lambda) = \mathbb{F}_{\text{cosine}}\{f(x)\} = \int_0^{\infty} f(x)\cos(\lambda x)dx$$

The function $f(x)$ is then called the *Inverse Fourier Cosine Transform* of $\mathbb{F}_{\text{cosine}}(\lambda)$. and is given by:

$$f(x) = \frac{2}{\pi}\int_0^{\infty}\mathbb{F}_{\text{cosine}}(\lambda)\cos(\lambda x)dx$$

Example 1:

(a) Find the Fourier transform of $f(x) = \begin{cases} 1 & |x| < a \\ 0 & |x| < a \end{cases}$.

(b) Graph $f(x)$ and its Fourier transform for $a = 1$.

The Fourier transform of $f(x)$ is

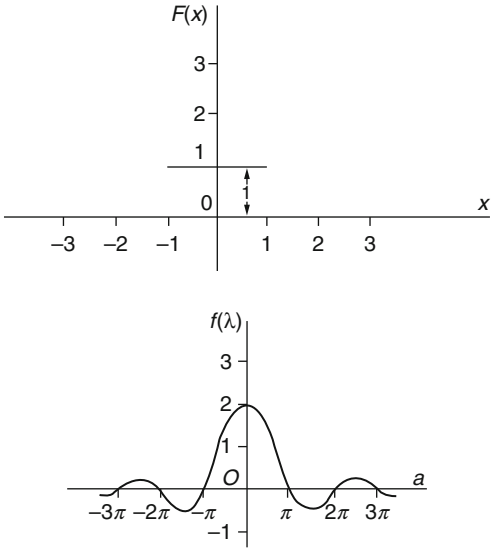
Solution: (a) The Fourier transform of $f(x)$ is

$$f(\lambda) = \int_{-\infty}^{\infty} f(u)e^{-i\lambda u}\quad du = \int_{-\infty}^{\infty}(1)e^{-i\lambda u}\quad du = \frac{e^{-i\lambda u}}{-i\lambda}\Big|_{-a}^{+a}$$

$$= \left(\frac{e^{i\lambda u} - e^{-i\lambda u}}{i\lambda}\right) \qquad = 2\frac{\sin\lambda a}{\lambda} \qquad \lambda \neq 0$$

For $\lambda = 0$, we obtain $f(\lambda) = 2a$.

(b) The graph of $f(x)$ and $f(\lambda)$ for $a = 1$ are shown down below.



Example 2: Use the result of above example (Example 1)

(a) To evaluate $\int_{-\infty}^{\infty} \frac{\sin \lambda a \cos \lambda x}{\lambda} d\lambda.$

(b) Deduce the value of $\int_{-\infty}^{\infty} \frac{\sin u}{u} du$

Solution: (a) From Fourier's integral theorem, if

$$f(\lambda) = \int_{-\infty}^{\infty} f(u) e^{-i\lambda u} du. \quad \text{then} \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(u) e^{-i\lambda u} d\lambda$$

Then from Example 1 above we have;

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin \lambda a}{\lambda} e^{i\lambda x} da = \begin{cases} 1 & |x| < a \\ 1/2 & |x| = a \\ 0 & |x| < a \end{cases} \tag{E.38}$$

The left side of Eq. (E.38) is equal to

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin \lambda a \cos \lambda x}{\lambda} da + \frac{i}{\pi} \int_{-\infty}^{\infty} \frac{\sin \lambda a \sin \lambda x}{\lambda} d\lambda \tag{E.39}$$

The integral in the second integral of Eq. (E.39) is odd and so the integral is zero. Then from Eqs. (E.38) and (E.39), we have

$$\int_{-\infty}^{\infty} \frac{\sin \lambda a \cos \lambda x}{\lambda} d\lambda = \begin{cases} \pi & |x| < a \\ \pi/2 & |x| = a \\ 0 & |x| < a \end{cases} \tag{E.40}$$

(b) If $x = 0$ and $a = 1$ in the result of (a), we have

$$\int_{-\infty}^{\infty} \frac{\sin \lambda}{\lambda} d\lambda = \pi \quad \text{or} \quad \int_{-\infty}^{\infty} \frac{\sin \lambda}{\lambda} d\lambda = \frac{\pi}{2}$$

Example 3: If $f(x)$ is an even function show that:

$$f(\lambda) = 2\int_{-\infty}^{\infty} f(u) \cos(\lambda u) du$$

$$f(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(\lambda) \cos(\lambda u) d\lambda$$

We have

$$f(\lambda) = \int_{-\infty}^{\infty} f(u) e^{-i\lambda u} du = \int_{-\infty}^{\infty} f(u) \cos(\lambda u) du - i\int_{-\infty}^{\infty} f(u) \sin(\lambda u) du \tag{E.41}$$

(a) If $f(u)$ is even, $f(u)\cos(\lambda u)$ is even and $f(u)\sin(\lambda u)$ is odd. Then the second integral on the right of Eq. (E.41) is zero and the result can be written

$$f(\lambda) = 2\int_{-\infty}^{\infty} f(u) \cos(\lambda u) du$$

(b) From (a), $f(-\lambda) = f(\lambda)$ so that $f(\lambda)$ is an even function. Then by using a proof exactly analogous to that in (a), the required result follows.

A similar result holds for odd functions and can be obtained by replacing the cosine by the sine.

E.18 The Convolution Theorem

The *convolution* of two functions $f(x)$ and $g(x)$, where $-\infty < x < \infty$, is defined as

$$f^*g = \int_{-\infty}^{\infty} f(u) g(x-u) du = h(x) \tag{E.42}$$

An important result, known as the *Convolution Theorem for Fourier Transforms*, is the following.

Theorem. *If* h(x) *is the convolution of* f(x) *and* g(x)*, then*

$$\int_{-\infty}^{\infty} h(x)e^{-i\lambda x}dx = \left\{\int_{-\infty}^{\infty} f(x)e^{-i\lambda x}dx\right\}\left\{\int_{-\infty}^{\infty} g(x)e^{-i\lambda x}dx\right\} \tag{E.43}$$

or

$$\mathbb{F}\{f^*g\} = \mathbb{F}\{f\}\mathbb{F}\{g\} \tag{E.44}$$

This says that the Fourier transform of the convolution of f and g is the product of the Fourier transform of f and g.

E.19 Parseval's Identity for Fourier Integrals

If the Fourier transform of $f(x)$ is $f(\lambda)$, then

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(\lambda)|^2 d\lambda \tag{E.45}$$

This is called *Parseval's identity for Fourier Integrals*. Generalizations of this are possible.

E.20 Relationships of Fourier and Laplace Transform

Consider the function

$$f(t) = \begin{cases} e^{-xt}\Phi(t) & t > 0 \\ 0 & t < 0 \end{cases} \tag{E.46}$$

Then from Eq. E.35, with λ replaced by y,we see that the Fourier transform of $f(t)$ is

$$\mathbb{F}\{f(t)\} = \int_0^{\infty} e^{-iyt}e^{-xt}\Phi(t)dt \quad = \int_0^{\infty} e^{-(x+iy)t}\Phi(t)dt \quad t > 0 \tag{E.47}$$

where we have written $s = x + iy$. The right of Eq. E.47 is the Laplace transform of $\Phi(t)$ and the result indicates a relationship of Fourier and Laplace transform. It also indicates a need for considering s as a complex variable $x + iy$.

The relationship can even be taken further and note that $f(t)$ and $g(t)$ are zero for $t < 0$, the convolution of f and g given by Eq. E.42 and can be written as follows:

$$f(t)*g(t) = \int_0^t f(u)g(t-u)du = h(x) \tag{E.48}$$

and Eq. E.44 corresponds to

$$\mathcal{L}\{f*g\} = \mathcal{L}\{f\}\mathcal{L}\{g\} \tag{E.49}$$

Example 1: Prove that for two function of $f(t)$ and $g(t)$ we have the relationship of $f(t)*g(t) = g(t)*f(t)$ for $t > 0$.

Solution: Letting $t - u = v$ or $u = t - v$ and using Eq. E.48 for $t > 0$, we have

$$f(t)*g(t) = \int_0^t f(u)g(t-u)du = \int_0^t f(t-v)g(v)dv$$

$$= \int_0^t f(v)g(t-v)dv = g(t)*f(t)$$

E.21 Summary of Fourier Transforms Series and Integrals

Let $f(x)$ satisfy the following conditions:
$f(x)$ is defined in the interval $c < x < c + 2L$.
$f(x)$ and $f'(x)$ are sectionally continuous in $c < x < c + 2L$.
$f(x + 2x) = f(x)$, i.e. $f(x)$ is periodic with period $2L$.
Then at every period of continuity, we have

$$f(x) = \frac{A_0}{2} + \sum_{n=1}^{\infty}\left[A_n \cos\left(\frac{n\pi x}{L}\right) + B_n \sin\left(\frac{n\pi x}{L}\right)\right] \tag{E.50}$$

where

$$\left.\begin{aligned} A_n &= \frac{1}{L}\int_c^{c+2L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx \\ B_n &= \frac{1}{L}\int_c^{c+2L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx \end{aligned}\right\} \tag{E.51}$$

(continued)

At a point of discontinuity, the left side of Eq. (E.51) is replaced by $\frac{1}{2}\{f(x+0)+f(x-0)\}$, i.e., the mean value at the discontinuity.

The series Eq. (E.50) with coefficients Eq. (E.51) is called the Fourier Series of $f(x)$. For many problems, $c = 0$ or $c = -L$. In case $L = \pi$,

Finite Fourier Sine Transform of $f(x), 0 < x < L$, is defined. $f(x)$ has period 2π and Eqs. E.3 and E.4 are simplified.

The above conditions are often called ***Dirichlet Conditions*** and are sufficient (but not necessary conditions for convergence of Fourier Series.

The *Finite Fourier Sine Transform* of $f(x)$, $0 < x < L$, is defined as.

$$\mathbb{F}_{\text{sine}}(n) = \mathbb{F}_{\text{sine}}\{f(x)\} = \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right)$$

The *Inverse Finite Fourier Sine Transform* of $\mathbb{F}_{\text{sine}}(n)$ for integer n.

$$f(x) = \frac{2}{L}\sum_{n=1}^{\infty} \mathbb{F}_{\text{sine}}(n) \sin\frac{n\pi x}{L}$$

The *Finite Fourier Cosine Transform* of $f(x)$, $0 < x < L$, is defined as.

$$\mathbb{F}_{\text{cosine}}(n) = \mathbb{F}_{\text{cosine}}\{f(x)\} = \int_0^L f(x) \cos\left(\frac{n\pi x}{L}\right)dx$$

The *Inverse Finite Fourier Cosine Transform* of $\mathbb{F}_{\text{cosine}}(n)$ for integer n.

$$f(x) = \frac{1}{L}\mathbb{F}_{\text{cosine}}(0) + \frac{2}{L}\sum_{n=1}^{\infty} \mathbb{F}_{\text{cosine}}(n) \cos\frac{n\pi x}{L}$$

The (*infinite*) Fourier Sine Transform of $f(x)$, $0 < x < \infty$, is defined as

$$\mathbb{F}_{\text{sine}}(\lambda) = \mathbb{F}_{\text{sine}}\{f(x)\} = \int_0^{\infty} f(x) \sin(\lambda x)dx$$

The function $f(x)$ is then called the *Inverse Fourier Sine Transform* of $\mathbb{F}_{\text{sine}}(\lambda)$ and is given by:

$$f(x) = \frac{2}{\pi}\int_0^{\infty} \mathbb{F}_{\text{sine}}(\lambda) \sin(\lambda x)dx$$

The (*infinite*) Fourier Cosine Transform of $f(x)$, $0 < x < \infty$, is defined as

(continued)

$$\mathbb{F}_{\text{cosine}}(\lambda) = \mathbb{F}_{\text{cosine}}\{f(x)\} = \int_0^\infty f(x)\cos(\lambda x)dx$$

The function $f(x)$ is then called the *Inverse Fourier Cosine Transform* of $\mathbb{F}_{\text{cosine}}(\lambda)$ and is given by:

$$f(x) = \frac{2}{\pi}\int_0^\infty \mathbb{F}_{\text{cosine}}(\lambda)\cos(\lambda x)dx$$

E.22 More Examples of Fourier Analysis

Example 1: Find (a) the Finite Fourier Sine Transform and (b) the Finite Fourier Cosine Transform of $\partial T(x,t)/\partial x$ where T is a function of x and t for $0 < x < L$, $t > 0$.

Solution: (a) By definition the finite Fourier sine transformation of $\partial T(x,t)/\partial x$ is, on integrating by parts.

$$\int_0^L \frac{\partial T(x,t)}{\partial x}\sin\left(\frac{n\pi x}{L}\right)dx = T(x,t)\sin\left(\frac{n\pi x}{L}\right)\Big|_0^L - \frac{n\pi}{L}\int_0^L T(x,t)\cos\left(\frac{n\pi x}{L}\right)dx$$

Second integral above is just the definition of Fourier Cosine Transform, while the first integral approaches zero for all the value of $n = 1, 2, 3, \ldots$. and interval of $0 < x < L$.

$$\mathbb{F}_{\text{sine}}\left\{\frac{\partial T(x,t)}{\partial x}\right\} = \int_0^L \frac{\partial T(x,t)}{\partial x}\sin\left(\frac{n\pi x}{L}\right)dx$$

$$= 0 - \frac{n\pi}{L}\int_0^L \frac{\partial T(x,t)}{\partial x}\sin\left(\frac{n\pi x}{L}\right)dx = -\frac{n\pi}{L}\mathbb{F}_{\text{cosine}}\{T(x,t)\}$$

(b) The finite Fourier cosine transform is

$$\int_0^L \frac{\partial T(x,t)}{\partial x}\cos\left(\frac{n\pi x}{L}\right)dx = T(x,t)\cos\left(\frac{n\pi x}{L}\right)\Big|_0^L + \frac{n\pi}{L}\int_0^L T(x,t)\sin\left(\frac{n\pi x}{L}\right)dx$$

First equation can be evaluated for the interval of $0 \le x \le L$ and second equation is just the definition of Fourier Sine Transform for all the value of $n = 1, 2, 3, \ldots$. and interval of $0 < x < L$.

$$\mathbb{F}_{\text{cosine}}\left\{\frac{\partial T(x,t)}{\partial x}\right\} = \int_0^L \frac{\partial T(x,t)}{\partial x}\cos\left(\frac{n\pi x}{L}\right)dx$$

$$= \frac{n\pi}{L}\int_0^L \frac{\partial T(x,t)}{\partial x}\sin\left(\frac{n\pi x}{L}\right)dx + \frac{n\pi}{L}\int_0^L T(x,t)\sin\left(\frac{n\pi x}{L}\right)dx$$

$$= \frac{n\pi}{L}\mathrm{F}_{\text{sine}}\{T(x,t)\} - \{T(0,t) - T(L,t)\cos n\pi\}$$

Example 2: Work problem (a) and (b) Example 1 in above for function $\dfrac{\partial^2 T(x,t)}{\partial x^2}$ where T is a function of x and t for $0 < x < L$, $t > 0$.

Solution: Replacing $T(x,t)$ by $\partial T(x,t)/\partial x$ in the result of Example 1 above

$$\mathbb{F}_{\text{sine}}\left\{\frac{\partial^2 T(x,t)}{\partial x^2}\right\} = \mathbb{F}_{\text{sine}}\left\{\frac{\partial^2 T}{\partial x^2}\right\} = -\frac{n\pi}{L}\mathbb{F}_{\text{cosine}}\left\{\frac{\partial T}{\partial x}\right\}$$

$$= -\frac{n^2\pi^2}{L^2}\mathrm{F}_{\text{sine}}\{T\} + \frac{n\pi}{L}\{T(0,t) - T(L,t)\cos n\pi\}$$

$$\mathbb{F}_{\text{cosine}}\left\{\frac{\partial^2 T(x,t)}{\partial x^2}\right\} = \mathbb{F}_{\text{sine}}\left\{\frac{\partial^2 T}{\partial x^2}\right\}$$

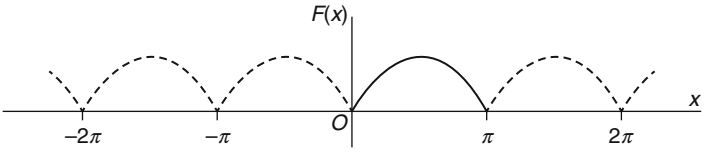
$$= -\frac{n\pi}{L}\mathbb{F}_{\text{sine}}\left\{\frac{\partial T}{\partial x}\right\} - \{T_x(0,t) - T_x(L,t)\cos n\pi\}$$

$$= -\frac{n^2\pi^2}{L^2}\mathbb{F}_{\text{sine}}\left\{\frac{\partial T}{\partial x}\right\} - \{T_x(0,t) - T_x(L,t)\cos n\pi\}$$

where T_x denotes the partial derivative with respect to x.

Example 3: Expand $f(x) = \sin x$ for interval $0 < x < \pi$, in a Fourier cosine series.

Solution: A Fourier series consisting of sine terms alone is obtained only for an even function. Hence we extend the definition of $f(x)$ so that it becomes even (dashed part of figure below). With this extension, $f(x)$ is defined in an interval of length 2π. Taking the period as 2π, we have $2L = 2\pi$ so that $L = \pi$.



By Example 1 of Sect. 1.3 of this appendix we have $B_n = 0$ and for A_n we have the following analysis;

$$A_n = \frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx = \frac{2}{\pi}\int_0^\pi \sin(x)\cos(nx)dx$$

$$= \frac{1}{\pi}\int_0^\pi \sin(x+\pi x) + \sin(x+\pi x)dx$$

$$= \tfrac{1}{\pi}\left\{-\frac{\cos(n+1)x}{n+1} - \frac{\cos(n-1)x}{n-1}\right\}\Bigg|_0^\pi$$

$$= \frac{1}{\pi}\left\{-\frac{1-\cos(n+1)\pi}{n+1} - \frac{\cos(n-1)\pi - 1}{n-1}\right\}$$

$$= \frac{1}{\pi}\left\{-\frac{1-\cos(n\pi)}{n+1} - \frac{\cos(n\pi)}{n-1}\right\}$$

$$= \frac{-2(1+\cos n\pi)}{\pi(n^2-1)} \text{ if } n \neq 1$$

For $n = 1$, $A_1 = \dfrac{2}{\pi}\displaystyle\int_0^\pi \sin x\cos x dx = \frac{2}{\pi}\frac{\sin^2 x}{2}\Big|_0^\pi = 0$

For $n = 0$, $A_0 = \dfrac{2}{\pi}\displaystyle\int_0^\pi \sin x dx = \frac{2}{\pi}(-\cos x)\Big|_0^\pi = \frac{\pi}{4}$

$$f(x) = \frac{2}{\pi} - \frac{2}{\pi}\sum_{n=2}^{\infty}\frac{(1+\cos n\pi)}{n^2-1}\cos nx$$

Then

$$\frac{2}{\pi} - \frac{4}{\pi}\left(\frac{\cos 2x}{2^2-1} + \frac{\cos 4x}{4^2-1} + \frac{\cos 6x}{6^2-1} + \ldots.\right)$$

Example 4: Show that $\displaystyle\int_0^\infty \frac{\cos(\lambda x)}{\lambda^2+1}d\lambda = \frac{\pi}{2}e^{-x}$ for $x \geq 0$.

Solution: Letting $f(x) = e^{-x}$ in the Fourier integral theorem we obtain the following;

$$f(x) = \frac{2}{\pi}\int_0^\infty \cos(\lambda x)d\lambda\int_0^\infty f(u)\cos(\lambda u)du$$

Then

$$\frac{2}{\pi}\int_0^\infty \cos(\lambda x)d\lambda\int_0^\infty e^{-u}\cos(\lambda u)du = e^{-x}$$

Since $\int_0^\infty e^{-u}\cos(\lambda u)du = \frac{1}{\lambda^2+1}$, we have

$$\frac{2}{\pi}\int_0^\infty \frac{\cos(\lambda x)}{\lambda^2+1}d\lambda = e^{-x} \quad \text{or} \quad \int_0^\infty \frac{\cos(\lambda x)}{\lambda^2+1}d\lambda = \frac{\pi}{2}e^{-x}$$

Example 5: Solve the integral equation $\int_0^\infty f(x)\cos(\lambda x)dx = \begin{cases} 1-\lambda & 0 \le \lambda \le 1 \\ 0 & \lambda > 1 \end{cases}$.

Solution: Let $\int_0^\infty f(x)\cos(\lambda x)dx = f(\lambda)$ and choose $f(\lambda) = \begin{cases} 1-\lambda & 0 \le \lambda \le 1 \\ 0 & \lambda > 1 \end{cases}$.
Then by Fourier's integral theorem,

$$\begin{aligned} f(x) &= \frac{2}{\pi}\int_0^\infty f(x)\cos(\lambda x)dxd\lambda \\ &= \frac{2}{\pi}\int_0^\infty (1-\lambda)\cos(\lambda x)dxd\lambda \\ &= \frac{2(1-\cos x)}{\pi x^2} \end{aligned}$$

E.23 Laplace Transformation

A very powerful technique for solving Ordinary and Partial Differential Equation problems is that of the Laplace transform, which literally transforms the original differential equation into an elementary algebraic expression. This latter can then simply be transformed once again, into the solution of the original problem. This technique is known as the "Laplace transform method". Solving these ODEs and PDEs with certain description and quantities vary with time such as the flow of heat through an insulated conductor. These equations are generally coupled with initial conditions that describe the state of the system at time $t = 0$.

Suppose that $f(t)$ is a real or complex variable function of the (time) variable t for $t > 0$ and s is real or complex parameter. We define the Laplace Transform of $f(t)$ as

$$\begin{aligned} f(s) = \pounds\{f(t)\} &= \int_0^\infty e^{-st} f(t)dt \\ &= \lim_{\tau\to\infty}\int_0^\tau e^{-st} f(t)dt \end{aligned} \tag{E.52}$$

whenever the limit exists (as a finite number). When it does, the integral Eq. E.52 is said to *converge*. If the limit does not exist, the integral is said to diverge and there is no Laplace transform for $f(t)$. The notation $\pounds\{f\}$ will also be used to denote the Laplace transform of f, and the integral is the ordinary Riemann (Improper) integral [4].

The parameter s belongs to some domain on the real line or in the complex plane. We will choose s appropriately so as to ensure the convergence of the Laplace integral (Eq. E.52). In a mathematical and technical sense, the domain of s is quite important. However, in a practical sense, when differential equations are solved, the domain of s is routinely ignored. When s is complex, we will always use the notation $s = x + iy$.

The symbol $\pounds$ is the Laplace transformation, which acts on functions $f = f(t)$ and generates a new function, $f(s) = \pounds\{f(t)\}$ [4].

Example 1: If $f(t) \equiv 1$ for $t \geq 0$, then

$$\begin{aligned} \pounds\{f(t)\} &= \int_0^\infty e^{-st} dt \\ &= \lim_{x\to\infty} \left(\frac{e^{-st}}{-s} \Big|_0^t \right) \\ &= \lim_{x\to\infty} \left(\frac{e^{-s\tau}}{-s} + \frac{1}{s} \right) \\ &\quad \frac{1}{s} \end{aligned} \tag{E.53}$$

provide of course that $s > 0$ (if s is real). Thus we have

$$\pounds\{1\} = \frac{1}{s} \quad (s > 0) \tag{E.54}$$

If $s \leq 0$, then the integral would diverge and there would be no resulting Laplace transform. If we had taken s to be a complex variable, the same calculation, with $\mathrm{Re}(s) > 0$, would have given $\pounds\{1\} = 1/s$. In fact, let us just verify that in the above calculation the integral can be treated in the same way even if s is a complex variable. We require the well-known Euler formula (see Appendix D)

$$e^{i\theta} = \cos\theta + i \sin\theta \tag{E.55}$$

and the fact that $\left|e^{i\theta}\right| = 1$. The claim is that (ignoring the minus sign as well as the limits of integration to simplify the calculation)

$$\int e^{st}dt = \frac{e^{st}}{s} \tag{E.56}$$

for $s = x + iy$ any complex number $\neq 0$. To see this observe that

$$\begin{aligned}\int e^{st}dt &= \int e^{(x+iy)t}dt \\ &= \int e^{xt}\cos(yt)dt + i\int e^{xt}\sin(yt)dt\end{aligned}$$

by Euler's formula. Performing a double integration by parts on both these integrals gives

$$\int e^{st}dt = \frac{e^{xt}}{x^2+y^2}[(x\cos(yt) + y\sin(yt) - y\cos(yt))]$$

Now the right-hand side of Eq.(E.56) can be expressed as

$$\begin{aligned}e^{st} &= \frac{e^{(x+iy)t}}{x+iy} \\ &= \frac{e^{xt}(\cos(yt) + i\sin(yt))(x-iy)}{x^2+y^2} \\ &= \frac{e^{xt}}{x^2+y^2}[(x\cos(yt) + y\sin(yt)) + i(x\sin(yt) - y\cos(yt))]\end{aligned}$$

which equals the left-hand side, and (Eq. E.56) follows.

Furthermore, we obtain the result of Eq.(E.54) for s complex if we take $\mathrm{Re}(s) = x > 0$, since then

$$\lim_{\tau\to\infty}|e^{-s\tau}| = \lim_{\tau\to\infty} e^{-x\tau} = 0$$

killing off the limit in Eq. E.54. Let us use the preceding to calculate $\pounds\{\cos(\omega t)\}$ and $\pounds\{\sin(\omega t)\}$ for ω (ω real).

Example 2: We begin with

$$\begin{aligned}\pounds\{e^{i\omega t}\} &= \int_0^\infty e^{st}e^{i\omega t}dt \\ &= \lim_{\tau\to\infty}\left.\frac{e^{(i\omega-s)t}dt}{i\omega - s}\right|_0^\tau \\ &= \frac{1}{s-i\omega}\end{aligned}$$

since $\lim_{\tau\to\infty} |e^{i\omega t}e^{-st}| = \lim_{\tau\to\infty} e^{-xt} = 0$, provided $x = \mathrm{Re}(s) > 0$. Similarly, $\pounds\{e^{-i\omega t}\} = 1/s + i\omega$. Therefore, using the linearity property of $\pounds$, which follows from the fact that integrals are linear operators.

$$\frac{\pounds(e^{i\omega t}) + \pounds(e^{-i\omega t})}{2} = \pounds\left(\frac{e^{i\omega t} + e^{-i\omega t}}{2}\right) = \pounds\{\cos(\omega t)\}$$

and consequently,

$$\pounds\{\cos(\omega t)\} = \frac{1}{2}\left(\frac{1}{s - i\omega} + \frac{1}{s - i\omega}\right) = \frac{1}{s^2 - \omega^2} \tag{E.57}$$

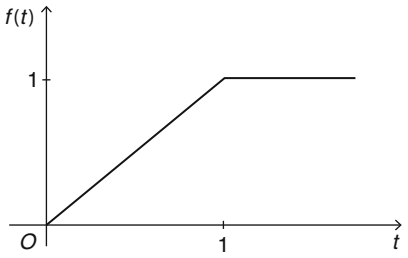
Similarly,

$$\pounds\{\sin(\omega t)\} = \frac{1}{2i}\left(\frac{1}{s - i\omega} + \frac{1}{s - i\omega}\right) = \frac{\omega}{s^2 - \omega^2} \quad (\mathrm{Re}(s) > 0) \tag{E.58}$$

The Laplace transform of functions defined in a piecewise fashion is readily handled as follows.

Example 3: Let figure below be

$$f(t) = \begin{cases} t & 0 \le \mathrm{t} \le 1 \\ 1 & t > 1 \end{cases}$$



From the definition

$$\begin{aligned} \pounds\{f(t)\} &= \int_0^\infty f(t)e^{-st}dt \\ &= \int_0^1 te^{-st}dt + \lim_{\tau\to\infty}\int_0^\tau e^{-st}dt \\ &= \frac{te^{-st}}{-s}\bigg|_0^1 + \frac{1}{s}\int_0^1 e^{-st}dt + \lim_{\tau\to\infty}\frac{e^{-st}}{-s}\bigg|_0^\tau \end{aligned}$$

Example 4: Let $f(t) = t$, then evaluate $\pounds\{f(t)\}$

Solution: From definition Eq. E.52 we have $\pounds\{f(t)\} = \int_0^\infty e^{-st} t dt$. Integrating by parts and using $\lim_{t\to\infty} te^{-st} = 0$ for $s > 0$, along with the result from Example 1, we obtain

$$\pounds\{t\} = \frac{-te^{-st}}{s}\bigg|_0^\infty + \frac{1}{s}\int_0^\infty e^{-st} dt = \frac{1}{s}\pounds\{t\} = \frac{1}{s}\left(\frac{1}{s}\right) = \frac{1}{s^2}$$

Example 5: Let $f(t) = \left\{e^{-3t}\right\}$, then evaluate $\pounds\{f(t)\}$

Solution: From definition Eq. E.52 we have

$$\pounds\{f(t)\} = \int_0^\infty e^{-st} e^{-3t} dt$$

or

$$\begin{aligned}\pounds\{f(t)\} = \int_0^\infty e^{-st} e^{-3t} dt \quad &= \int_0^\infty e^{-(s+3)t} dt \\ &= \frac{e^{-(s+3)t}}{s+3}\bigg|_0^\infty \\ &= \frac{1}{s+3} \qquad\qquad s > -3\end{aligned}$$

E.24 Laplace Transform is a Linear Transform

For a linear combination of functions, we can write

$$\int_0^\infty e^{-st}[\alpha f(t) + \beta g(t)]dt = \alpha\int_0^\infty e^{-st} f(t)dt + \beta\int_0^\infty e^{-st} g(t)dt$$

whenever both integrals converge for $s > c$. Hence it follows that

$$\pounds[\alpha f(t) + \beta g(t)] = \alpha\pounds\{f(t)\} + \beta\pounds\{f(t)\} = \alpha f(s) + \beta g(s) \tag{E.59}$$

Because of the property give in Eq. E.59, $\pounds$ operator is said to be a **Linear Transform**. For example, from Examples of 1 and 4, we can write the following,

$$£\{1+t\} = £\{1\} + £\{t\} = \frac{1}{s} + \frac{1}{s^2}$$

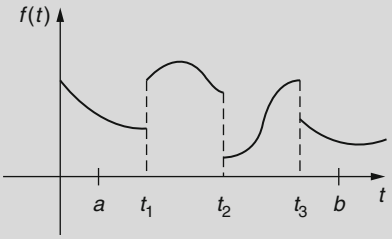
We state the generalization of some of the preceding examples by means of the next theorem. It is understood that s is sufficiently restricted to guarantee the convergence of the appropriate Laplace $\{e^{at}\} = \dfrac{1}{s-a}$ transformation [7].

Theorem: Transforms of Some Basic Functions [7]

(a) $£\{1\} = \dfrac{1}{s}$	
(b) $£\{t^n\} = \dfrac{n!}{s^{n+1}} \quad n = 1, 2, 3 \ldots$	(c) $£\{e^{at}\} = \dfrac{1}{s}$
(d) $£\{\sin(\omega t)\} = \dfrac{\omega}{s^2+\omega^2}$	(e) $£\{\cos(\omega t)\} = \dfrac{s}{s^2+\omega^2}$
(f) $£\{\sinh(\omega t)\} = \dfrac{\omega}{s^2+\omega^2}$	(g) $£\{\cosh(\omega t)\} = \dfrac{s}{s^2+\omega^2}$

Sufficient Conditions for Existence of Laplace Transform £{*f*(*t*)} [7]
The integral that defines the Laplace transform does not have to converge. For example, neither $£\{f(1/t)\}$ nor $£\left\{f\left(e^{t^2}\right)\right\}$ exists. Sufficient conditions guaranteeing the existence of $£\{f(t)\}$ are that f be piecewise continuous on $[0, \infty)$ and that f be of exponential order for $t > T$.

Piecewise Continuous Definition [7]
A function f is piecewise continuous on $[0, \infty)$ if, in any interval $0 \le a \le t \le b$, there are at most a finite number of points t_k, $k = 1, 2, \ldots, n$ for $(t_{k-1} < t_k)$ at which f has finite discontinuities and is continuous on each open interval $t_{k-1} < t < t_k$. See Figure below.



The concept of Exponential order is defined in the following manner.

Exponential Order
A function f is said to be of **exponential order** if there exist constants $£\{f^{\cdot}\}$, $M > 0$, and $T > 0$ such that $|f(t)| \le Me^{ct}$ all $t > T$.

For more details reader should refer to any Laplace Transform book or references at the end of this appendix

E.25 Derivative Theorem

In order to solve differential equations, it is necessary to know the Laplace transform of the derivative $f'(t) = f'$ of a function $f(t) = f$. The virtue of $\pounds\{f'\}$ is that it can be written in terms of $\pounds\{f(t)\}$.

Derivative Theorem
Suppose that f is continuous on $(0, \infty)$ and of exponential order α and that f' is piecewise continuous on $(0, \infty)$. Then

$$\pounds\left\{f'(t)\right\} = s\pounds\{f(t)\} - f(0)$$

As pointed out in any Laplace Transform book or differential equations both Ordinary Differential Equation (ODE) and Partial Differential Equation (PDE), the immediate goal is to use the Laplace transform to solve these types of differential equations. To that end we need to evaluate quantities such as $\pounds\{dy/dt\}$ and $\pounds\{d^2y/dt^2\}$. For example, if f' is continuous for $t \geq 0$ then integration by parts gives

$$\begin{aligned}\pounds\{f'(t)\} &= \int_0^\infty e^{-st} f'(t)dt = e^{-st}f(t)\Big|_0^\infty e^{-st}f(t)dt \\ &= -f(0) + s\pounds\{f(t)\}\end{aligned}$$

or

$$\pounds\left\{f'(t)\right\} = s\pounds\{f(t)\} - f(0) \tag{E.60}$$

Here we have assumed that $e^{-st}f(t) \to 0$ as $t \to \infty$. Similarly, with the aid of Eq. E.60,

$$\begin{aligned}\pounds\{f''(t)\} &= \int_0^\infty e^{-st} f''(t)dt = e^{-st}f'(t)\Big|_0^\infty + s\int_0^\infty e^{-st}f'(t)dt \\ &= -f'(0) + s\pounds\{f'(t)\} \\ &= s[sf(s) - f(0)] - f'(0)\end{aligned}$$

or

$$\pounds\{f''(t)\} = s^2 f(s) - sf(0) - f'(0) \tag{E.61}$$

In like manner it can be shown [7].

$$\pounds\{f''(t)\} = s^3 f(s) - sf^2(0) - sf'(0) - f''(0) \tag{E.62}$$

The recursive transform of the derivatives of a function f should be apparent from the results in Eqs. E.60, E.61 and E.62. The next theorem gives the Laplace transform of the n^{th} derivative of f. The proof is omitted [7].

Theorem: Transform of a Derivative
If $f, f', \ldots, f^{(n-1)}$ are continuous on $[(0, \infty)$ and are of exponential order and if is piecewise continuous on , then

$$\pounds\{f^n(t)\} = s^n f(s) - s^{n-1} f(0) - s^{n-1} f(0) - s^{n-2} - \cdots - f^{(n-1)}$$

where $f(s) = \pounds\{f(t)\}$.

We now utilize a theorem know as Translation Theorem to present two very useful results for determing Laplace transform and their inverses. The first pertains to a translation in the s-domain and the second to a translation in the t-domain [4].

First Translation Theorem
If $f(s) = \pounds\{f(t)\}$ for $\mathrm{Re}(s) > 0$, then
$f(s-a) = \pounds\{e^{at} f(t)\}$ {real, $\mathrm{Re}(s) > a$}

Proof:
For $\mathrm{Re}(s) > a$,

$$\begin{aligned} f(s-a) &= \int_0^\infty e^{-(s-a)t} f(t)dt \\ &= \int_0^\infty e^{-st} e^{at} f(t)dt \\ &= \pounds\{e^{at} f(t)\} \end{aligned}$$

Example 1: Determine $f(s) = £\{f(t)\}$ for $f(t) = te^{at}$ and $t > 0$.

Solution: Since

$$£\{t\} = \frac{1}{s} \quad (\mathrm{Re}(s) > 0)$$

Then

$$£\{ta^{at}\} = \frac{1}{(s-a)^2} \quad \mathrm{Re}(s) > 0$$

and in general

$$£\{t^n e^{at}\} = \frac{n!}{(s-a)^{n+1}} \quad n = 0, 1, 2, \ldots \quad \mathrm{Re}(s) > 0$$

This gives a useful inverse

$$£^{-1}\left\{\frac{n!}{(s-a)^{n+1}}\right\} = \frac{1}{n!} t^n e^{at}$$

Theorem: Translation Theorem Transforms of Some Basic Functions [4]

Since $£\{\sin(\omega t)\} = \dfrac{\omega}{s^2 + \omega^2}$ then

$£\{e^{2t}\sin(\omega t)\} = \dfrac{3}{(s-2)^2 + 3^2} = \dfrac{3}{(s-2)^2 + 9}$.

In general, we can define for $(\mathrm{Re}(s) > a)$ in all below cases.

(a) $£\{e^{at}\sin(\omega t)\} = \dfrac{\omega}{(s-a)^2 + \omega^2}$	(b) $£\{e^{at}\cos(\omega t)\} = \dfrac{s-a}{(s-a)^2 + \omega^2}$
(c) $£\{e^{at}\cosh(\omega t)\} = \dfrac{s-a}{(s-a)^2 + \omega^2}$	(d) $£\{e^{at}\sinh(\omega t)\} = \dfrac{\omega}{(s-a)^2 + \omega^2}$

Example 2: Determine the Laplace transform of the Laguerre polynomials, defined by

$$L_n(t) = \frac{e^t}{n!}\frac{d^n}{dt^n}(t^n e^{-t}) \quad n = 0, 1, 2 \ldots.$$

Solution: Let $y(t) = t^n e^{-t}$. Then

$$\pounds\{L_n(t)\} = \pounds\left(e^{-t}\frac{1}{n!}y^{(n)}\right)$$

First, we find by theorem of Transform of a Derivative in above and subsequently the first translation theorem in above as well coupled with Example 1 here,

$$\pounds\left\{y^{(n)}\right\} = s^n\pounds\{y\} = \frac{s^n n!}{(s+1)^{n+1}}$$

It follows that

$$\pounds\{L_n(t)\} = \pounds\left\{e^t\frac{1}{n!}y^{(n)}\right\} = \frac{(s-1)^n}{(s+1)^{n+1}} \quad (\mathrm{Re}(s) > 1)$$

again by the first translation theorem.

Second Translation Theorem

$$f(s) = \pounds\{f(t)\} \quad \text{for} \quad \mathrm{Re}(s) > 0, \text{ then}$$

$\pounds\{u_a(t)f(t-a)\} = e^{-as}f(s) \quad a \geq 0$

This followins from the basic fact that

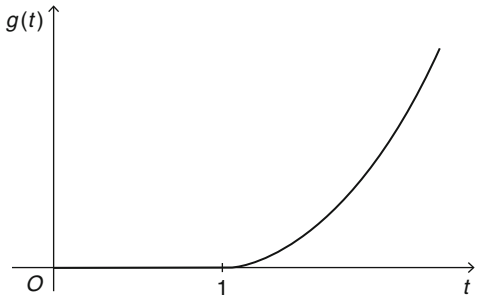
$$\int_0^\infty e^{-st}[u_a(t)f(t-a)]dt = \int_0^\infty e^{-st}f(t-a)dt$$

and setting $\tau = t - a$, the right-hand integral becomes

$$\int_a^\infty e^{-s(\tau+a)}f(\tau)d\tau = e^{-as}\int_a^\infty e^{-st}f(\tau)dt$$

Example 3: Let us determine $\pounds\{g(t)\}$ for the following figure.

$$g(t) = \begin{cases} 0 & 0 \leq t \leq 1 \\ (t-1)^2 & t \geq 1 \end{cases}$$



Note that $g(t)$ is just the function$f(t) = t^2$ delayed by $(a = 1)$ unit of time. Hence

$$\begin{aligned}\mathcal{L}\{g(t)\} &= \mathcal{L}\left\{u_1(t)(t-1)^2\right\} \\ &= e^{-s}\mathcal{L}\left\{t^2\right\} \qquad (\mathrm{Re}(s) > 0) \\ &= \frac{2e^{-s}}{s^3}\end{aligned}$$

Theorem: Translation Theorem Transforms of Some Special Functions [4]

Dirac Delta Function

For $t_0 > 0$, $\mathcal{L}[\delta(t - t_0)] = e^{-st_0}$

Beta Function

If $f(t) = t^{a-1}$ and $g(t) = t^{b-1}$ with $a, b > 0$ then $\mathcal{L}\left\{t^{a+b-1}B(a,b)\right\} = \mathcal{L}\left\{t^{a-1}\right\}\mathcal{L}\left\{t^{b-1}\right\} = \dfrac{\Gamma(a)(b)}{s^{a+b}}$

Error Function

For given $t > 0$ $\mathcal{L}\left\{erf\left(\sqrt{t}\right)\right\} = \frac{1}{s\sqrt{s+1}}$

Gamma Function

Let $f(t) = t^p$, with $p \leq 0$. Set $F(s) = \mathcal{L}\{f(t)\}$. It is easy to see that $F(s)$ is defined for $s > 0$. Consider the new variable $r = st$. Then we have

$$\mathcal{L}\{f(t)\} = \int_0^\infty e^{-st}t^p dt = \frac{1}{s^{p+1}}\int_0^\infty e^{-r}r^p dr$$

Which implies

$$\mathcal{L}\{t^p\} = \frac{\Gamma(p+1)}{s^{p+1}}$$

where the Gamma function $\Gamma(x)$ is defined by

(continued)

$$\Gamma(x) = \int_0^\infty e^{-r} r^{x-1} dr$$

The domain of the Gamma function is $x > 1$. Using integration by part, one can easily prove the fundamental formula

$$\Gamma(x+1) = x\Gamma(x) \quad x > 1$$

which implies (knowing that $\Gamma(1) = 1$)

$$\Gamma(n+1) = n! \quad n - 1, 2, 3, \ldots$$

Consequently, we have

$$\pounds\{t^n\} = \frac{n!}{s^{n+1}} \quad n - 1, 2, 3, \ldots$$

E.26 The Inverse Laplace Transform

If $f(s)$ represents the Laplace transform of a function $f(t)$—that is, $\pounds\{f(t)\} = f(s)$—we then say that $f(t)$ is the **Inverse Laplace Transform** of $f(s)$ and write $f(t) = \pounds^{-1}\{f(s)\}$. For example, from Example 1 and Example 2 of Sect. 2.0 above have respectively;

$$1 = \pounds^{-1}\left\{\frac{1}{s}\right\} \quad t = \pounds^{-1}\left\{\frac{1}{s^2}\right\} \text{ and } e^{-3t} = \pounds^{-1}\left\{\frac{1}{s+3}\right\}$$

The analogous of **Theorem: Transforms of Some Basic Functions** in Sect. 2 here for the inverse transform is presented as follows:

Theorem: Inverse Transform of Some Basic Functions [7]:

(a) $1 = \pounds^{-1}\left\{\frac{1}{s}\right\}$	
(b) $t^n = \pounds^{-1}\left\{\frac{n!}{s^{n+1}}\right\} \quad n = 1, 2, 3, \ldots$	(c) $e^{at} = \pounds^{-1}\left\{\frac{1}{s-a}\right\}$
(d) $\sin(\omega t) = \pounds^{-1}\left\{\frac{\omega}{s^2+\omega}\right\}$	(e) $\cos(\omega t) = \pounds^{-1}\left\{\frac{s}{s^2+\omega^2}\right\}$
(f) $\sinh(\omega t) = \pounds^{-1}\left\{\frac{\omega}{s^2+\omega}\right\}$	(g) $\cosh(\omega t) = \pounds^{-1}\left\{\frac{s}{s^2-\omega^2}\right\}$

When evaluating inverse transforms, it often happens that a function of s under consideration does not match exactly the form of a Laplace transform $f(s)$ given in the above table. It may be necessary to "fix up" the function of s by multiplying and dividing by an appropriate constant [7].

Example 1: Evaluate **(a)** $\pounds^{-1}\left\{\frac{1}{s^5}\right\}$ **(b)** $\pounds^{-1}\left\{\frac{1}{s^2+4}\right\}$

Solution: To match the form given in part (b) above table, we identify $n+1=5$ or $n=4$ and then multiply and dived by 4 !;

$$\pounds^{-1}\left\{\frac{1}{s^5}\right\} = \frac{1}{4!}\pounds^{-1}\left\{\frac{4!}{s^5}\right\} = \frac{1}{24}t^4$$

To match the form given in part (d) of above table, we identify $k^2=8$ and so $k=2$. We fix up the expression by multiplying and dividing by 2 then we have:

$$\pounds^{-1}\left\{\frac{1}{s^2+4}\right\} = \frac{1}{2}\pounds^{-1}\left\{\frac{1}{s^2+4}\right\} = \frac{1}{2}\sin(2t)$$

E.27 The Inverse Laplace is a Linear Transform

The inverse Laplace transform is also a linear transform; that is, for constants α and β.

$$\pounds^{-1}\{af(s)+\beta g(s)\} = a\pounds^{-1}\{f(s)\} + \beta\pounds^{-1}\{g(s)\} \tag{E.63}$$

where f and g are the transforms of some functions f and g, which extends to any finite linear combination of Laplace transforms.

Example 1: Evaluate $\pounds^{-1}\left\{\frac{-2s+6}{s^2+4}\right\}$

Solution: We first rewrite the given function of s as two expressions by means of term-wise division and then use Eq. E.67 and steps in Example 1 in above, we have:

$$\begin{aligned}\pounds^{-1}\left\{\frac{-2s+6}{s^2+4}\right\} &= \pounds^{-1}\left\{\frac{-2s}{s^2+} + \frac{6}{s^2+4}\right\} \\ &= -2\pounds^{-1}\left\{\frac{s}{s^2+4}\right\} + 6\pounds^{-1}\left\{\frac{6}{s^2+4}\right\} \\ &= -2\cos 2t + 3\sin 2t\end{aligned}$$

Example 2: Using the partial fraction technique from calculus, evaluate $\pounds^{-1}\left\{\dfrac{s^2 6s+9}{(s-1)(s-2)(s+4)}\right\}$.

Solution: There exist unique real constants A, B and C so that

$$\begin{aligned}\frac{s^2+6s+9}{(s-1)(s-2)(s+4)} &= \frac{A}{s-1}+\frac{B}{s-2}+\frac{C}{s+4}\\ &= \frac{A(s-2)(s+4)+B(s-1)(s+4)+C(s-1)(s-2)}{(s-1)(s-2)(s+4)}\end{aligned}$$

Since the denominators are identical, the numerators are identical, therefore

$$s^2+6s+9 = A(s-2)(s+4)+B(s-1)(s+4)+C(s-1)(s-2) \qquad \text{(E.64)}$$

By comparing coefficients of powers of s on both sides of the equality, we know that Eq. (E.64) is equivalent to a system of three equations in the three unknowns A, B and C. However, we can do the following analysis and we will find that:

$$\begin{aligned}s^2+6s+9 &= A(s-2)(s+4)+B(s-1)(s+4)+C(s-1)(s-2)\\ &= (A+B+C)s^2+(2A+3B-3C)s+(-8A-4B+2C)\end{aligned}$$

$$\begin{cases} A+B+C=1 \\ 2A+3B-3C \\ -8A-4B+2C=9 \end{cases}$$

$$\begin{cases} A=-\dfrac{16}{5} \\ B=\dfrac{25}{6} \\ C=\dfrac{1}{30} \end{cases}$$

Then we get

$$\frac{s^2+6s+9}{(s-1)(s-2)(s+4)} = \frac{16/5}{s-1}+\frac{25/6}{s-2}+\frac{1/30}{s+4} \qquad \text{(E.65)}$$

and thus, from the linearity of $\pounds^{-1}$ and the fact that $e^{at}=\pounds^{-1}\left\{\dfrac{1}{s-a}\right\}$ we have:

$$\pounds^{-1}\left\{\frac{s^2+6s+9}{(s-1)(s-2)(s+4)}\right\}=-\frac{16}{5}\pounds^{-1}\left\{\frac{1}{s-1}\right\}+\frac{25}{6}\pounds^{-1}\left\{\frac{1}{s-2}\right\}+\frac{1}{30}\pounds^{-1}\left\{\frac{1}{s+4}\right\}$$
$$=-\frac{16}{5}e^{t}+\frac{25}{6}e^{2t}+\frac{1}{30}e^{-4t} \tag{E.66}$$

E.28 Solving Linear Ordinary Differential Equations using Laplace Transform

It is apparent from the general result given in above (Theorem of Transform of a Derivative) that $\pounds\{d^n y/dt^n\}$ depends on $y(s)=\pounds\{y(t)\}$ and the $n-1$ derivative of $y(t)$ evaluated at $t=0$. This property makes the Laplace transform ideally suited for solving linear initial-value problems in which the differential equation has *constant coefficients*. Such a differential equation is simply a linear combination of terms $y, y', y'', \ldots. y^{(n)}$:

$$a_n\frac{d^n y}{dt^n}+a_{n-1}\frac{d^{n-1y}}{dt^{n-1}}+\cdots\cdot+a_0 y=g(t)$$

$$y(0)=y_0, y^{'}(0)=y_1 y^{''}(0)=y_2, \ldots., y^{(n-1)}(0)=y_{n-1}$$

where the a_i, $i=0,1,\cdots\cdot,n$ and $y_0, y_1, y_1, \cdots\cdot, y_{n-1}$ are constants. By the linearity property, the Laplace transform of this linear combination is a linear combination of Laplace transforms [7]:

$$a_n\pounds\left\{\frac{d^n y}{dt^n}\right\}+a_{n-1}\pounds\left\{\frac{d^{n-1}y}{dt^{n-1}}\right\}+\cdots\cdot+a_0\pounds\{g(t)\} \tag{E.67}$$

From the above theorem and Eq. E.67 becomes

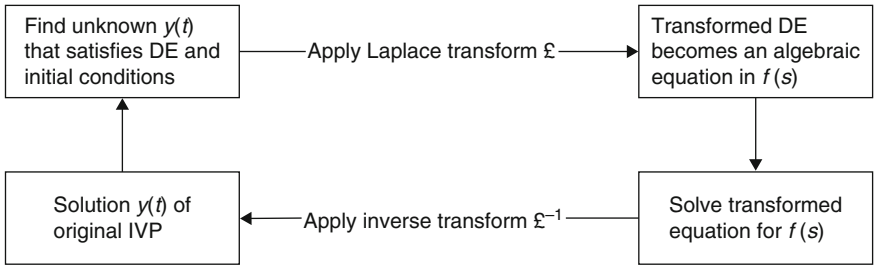
$$a_n\left[s^n y(s)-s^{n-1}y(0)-\cdots\cdot-y^{(n-1)}(0)\right]$$
$$+a_{n-1}\left[s^{n-1}y(s)-s^{n-2}y(0)-\cdots\cdot-y^{(n-2)}(0)\right]+\cdots\cdot+a_0 y(s)=g(s) \tag{E.68}$$

where $\pounds\{y(t)\}=y(s)$ and $\pounds\{g(t)\}=g(s)$. In other words, *the Laplace Transform of a linear differential equation with constant coefficient becomes an algebraic equation in* $y(s)$. If we solve the general transformed Eq. E.68 for the symbol $y(s)$, we first obtain $p(s)y(s)=q(s)+g(s)$, and then write,

$$y(s) = \frac{q(s)}{p(s)} + \frac{g(s)}{p(s)} \tag{E.69}$$

where $p(s) = a_0 s^n + a_{n-1} s^{n-1} + \cdots + a_0$, $q(s)$ is a polynomial in s of degree less than or equal to $n-1$ consisting of the various products of the coefficient a_i, $i = 0, 1, \cdots, n$ and the prescribed initial condition $y_0, y_1, \cdots, y_{n-1}$, and $g(s)$ is the Laplace transform of $g(t)$. Typically as we put decompose the expression into two or more partial fractions. Finally, the solution $y(t)$ of the original initial-value problem is $y(t) = \pounds^{-1}\{y(s)\}$.

The procedure is summarized in the diagram [7].


Find unknown y(t) that satisfies DE and initial conditions
Apply Laplace transform £
Transformed DE becomes an algebraic equation in f(s)
Solution y(t) of original IVP
Apply inverse transform £–1
Solve transformed equation for f(s)

E.29 The Complex Inversion Formula

In Appendix D we extensively showed the Complex Variable Analysis as well as the Complex Inversion Formula and we briefly explain it here again [3].

If $f(s) = \pounds\{f(t)\}$, then $\pounds^{-1}\{f(s)\}$ is given by

$$\begin{aligned} f(t) &= \pounds^{-1}\{f(s)\} \\ &= \frac{1}{2\pi i} \lim_{T\to\infty} \int_{\gamma - iT}^{\gamma + it} e^{st} f(s) ds = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{st} f(s) ds \quad t > 0 \end{aligned} \tag{E.70}$$

and $f(t) = 0$ for $t > 0$. This result is called the *Complex Inversion Integral* or *Formula*. It is also known as *Bromwich's integral formula*. The result provides a direct means for obtaining the inverse Laplace transform of given function $f(s)$.

The integration in Eq. E.70 is to be performed along a line $s = \gamma$ in the complex plane where $s = x + iy$. The real number γ is chosen so that $s = \gamma$ lies to the right of all the singularities (poles, branch points or essential singularities) but is otherwise arbitrary [3].

Example 1: Establish the validity of the complex inversion formula

Solution: We have, by definition: $f(s) = \int_0^\infty e^{-su} f(u)du$. Then

$$\lim_{T\to\infty} \frac{1}{2\pi i} \int_{\gamma-iT}^{\gamma+iT} e^{st} f(s)ds = \lim_{T\to\infty} \frac{1}{2\pi i} \int_{\gamma-iT}^{\gamma+iT} \int_0^\infty e^{-su} f(u)duds$$

Letting $s = \gamma + iy$, $ds = idy$, this becomes

$$\lim_{T\to\infty} \frac{1}{2\pi} e^{\gamma t} \int_{-T}^{T} e^{iyt} dy \int_0^\infty e^{-iyu} [e^{-\gamma u} f(u)]du = \frac{1}{2\pi} e^{\gamma t} \begin{cases} 2\pi e^{-\gamma t} f(t) & t > 0 \\ 0 & t < 0 \end{cases}$$

$$= \begin{cases} f(t) & t > 0 \\ 0 & t < 0 \end{cases}$$

By Fourier's integral theorem (See Sect. 1.13 of this appendix). Thus we have

$$f(t) = \tfrac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} f(s)ds \quad \text{for} \quad t > 0$$

as required.

Note: In the above proof, the assumption is that $e^{-\gamma u} f(u)$ is absolutely integrable in $(0, \infty)$, i.e. $\int_0^\infty e^{-\gamma u} |f(u)|du$ converges, so that Fourier's integral theorem can be applied. To insure this condition it is sufficient that $f(t)$ be of exponential order γ where the real number γ is chosen so that the line $s = \gamma$ in the complex plane lies to the right of all the singularities of $f(s)$. Except for this condition, γ is otherwise arbitrary.

E.30 The Bromwich Contour

In practice, the integral Eq. E.70 is evaluated by considering the contour integral Eq. E.71 as follows:

$$\frac{1}{2\pi i} \oint_C e^{-st} F(s)ds \tag{E.71}$$

where C is the contour of Fig. E.4. This contour, sometimes called the *Bromwich Contour*, is composed of the AB and the arc $BJKLA$ of a circle of radius R with center at the origin O. If we represent arc $BJKLA$ by Γ, it follows from Eq. E.70 that since $T = \sqrt{R^2 - \gamma^2}$,

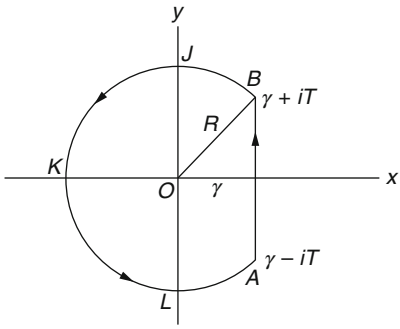


Fig. E.5 Different Functions Configurations

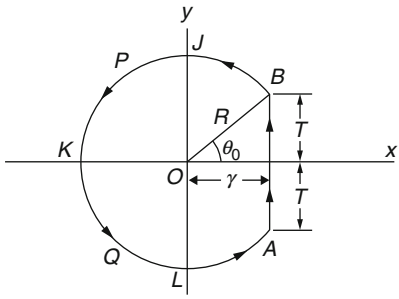


Fig. E.6 Different Functions Configurations

$$f(t) = \lim_{R\to\infty} \frac{1}{2\pi i} \int_{\gamma-iT}^{\gamma+iT} e^{st} f(s)ds = \lim_{R\to\infty} \left\{ \frac{1}{2\pi i} \oint_C e^{st} f(s)ds - \frac{1}{2\pi i} \int_{\Gamma} e^{st} f(s)ds \right\} \quad \text{(E.72)}$$

Example 1: Let Γ denote the curved portion $BJPKQLA$ of the Bromwich contour (See Fig. E.5 below) with equation $s = R^{i\theta}$, $\theta_0 \leq \theta \leq 2\pi - \theta_0$, i.e. Γ is the arc of a circle of radius R with center at O (Fig. E.6).

Solution: Suppose that on Γ we have

$$|f(s)| < \frac{M}{R^k}$$

where $k > 0$ and M are constants. Show that

$$\lim_{R\to\infty} \int_{\Gamma} e^{st} f(s)ds = 0$$

If Γ_1, Γ_2, Γ_3 and Γ_4 represent arcs BJ, JPK, KQL and LA respectively, we have

$$\int_{\Gamma} e^{st} f(s)ds = \int_{\Gamma_1} f(s)ds + \int_{\Gamma_2} e^{st} f(s)ds + \int_{\Gamma_3} e^{st} f(s)ds + \int_{\Gamma_4} e^{st} f(s)ds$$

$$= \int_{BJ} e^{st} f(s)ds + \int_{JPK} e^{st} f(s)ds + \int_{KQL} e^{st} f(s)ds + \int_{LA} e^{st} f(s)ds$$

What we need to show here that all the integrals on right hand side approach zero as $R \to \infty$, then we have proved the required result as below.

$$\lim_{R\to\infty} \int_{\Gamma} e^{st} f(s)ds = 0$$

Details of these analyses can be found in [4] and [7] of this appendix where readers should look it if they are interested [4, 7].

E.31 Use of Residue in Finding Inverse Laplace Transforms

Suppose that the only singularities of $f(s)$ are poles all of which lie to the left of the line $s = \gamma$ for some real constant γ. Suppose [3] further that the integral around Γ in Equation E-46 approaches zero as $R \to \infty$. Then by the residue theorem (See Sect. 30, Appendix D) we can write Eq. E.72 as;

$$\begin{aligned} f(t) &= \text{sum of residues of } e^{st} f(s) \text{ at poles of } f(s) \\ &= \sum \text{residues of } e^{st} f(s) \text{ at poles of } f(s) \end{aligned} \tag{E.73}$$

Example 1: Suppose that the only singularities of $f(s)$ are poles which all lie to the left of the line $s = \gamma$ for some real constant γ. Suppose further that $f(s)$ satisfies the condition given in Example 1 in pervious section. Prove that the inverse Laplace transform of $f(s)$ is given by

$f(t)$ =sum of residues of $e^{st} f(s)$ at all the poles of $f(s)$

Solution: We have

$$\frac{1}{2\pi i} \oint_C e^{st} f(s)ds = \frac{1}{2\pi i} \int_{\gamma - iT}^{\gamma + iT} e^{st} f(s)ds + \frac{1}{2\pi i} \int_{\Gamma} e^{st} f(s)ds$$

Where C is the Bromwich contour of Example 1 in above and Γ is the circular arc *BJPKQLA* of Fig. E.7. By residue theorem,

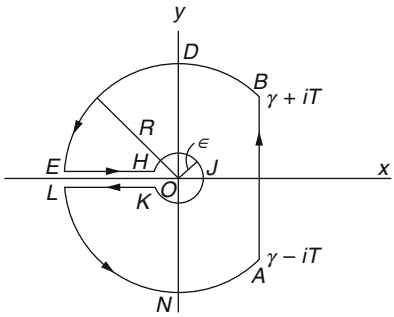


Fig. E.7 Different Functions Configurations

$\frac{1}{2\pi i}\oint_C e^{st}f(s)ds =$ sum of residues of $e^{st}f(s)$ at all poles of $f(s)$ inside $C = \sum$ residues inside C.

Thus $\frac{1}{2\pi i}\int_{\gamma-iT}^{\gamma+iT} e^{st}f(s)ds = \sum \text{residues inside C-}\frac{1}{2\pi i}\int_{\Gamma} e^{st}f(s)ds$

Taking the limit as $R \to \infty$, we find by Example 1 in previous section,

$$f(t) = \text{sum of residues of } e^{st}f(s) \text{ at all the poles } f(s)$$

Example 2: Show that $f(s) = \dfrac{1}{s-2}$ satisfies the condition in Example 1 of Sect. 2.7.

Find the residue of $\dfrac{e^{st}}{s-2}$ at the pole $s = 2$.

Evaluate $\pounds^{-1}\left\{\dfrac{1}{s-2}\right\}$ by using the complex inversion formula.

Solution: (a) For $s = \text{Re}^{i\theta}$, we have

$$\left|\frac{1}{s-2}\right| = \left|\frac{1}{\text{Re}^{i\theta}-2}\right| \le \frac{1}{|\text{Re}^{i\theta}|-2} = \frac{1}{R-2} < \frac{2}{R}$$

for large enough R (e.g. $R > 4$). Thus the condition in Example 1 of Sect. 2.7 is satisfied when $k = 1, M = 2$. Note that in establishing the above we have used the result $|z_1 - z_2| \le |z_1| - |z_2|$.

(b) The residue at the simple pole $s = 2$ is

$$\lim_{s\to 2}(s-2)\left(\frac{e^{st}}{s-2}\right) = e^{2t}$$

(c) By Example 1 Sect. 2.8 and the results of part (a) and (b), we see that

$\pounds^{-1}\left\{\dfrac{1}{s-2}\right\} =$ sum of residues of $e^{st}f(s) = e^{2t}$

Note that the Bromwich contour in this case is chosen so that is any real number greater than 2 and the contour encloses the pole $s = 2$.

Example 3: Find the residues of $f(s) = \dfrac{1}{s(s-a)}$

Solution: $f(s)$ has a simple pole at $s = 0$ and $s = a$, and $|f(s)| \le M/|s|^2$ for all $|s|$ sufficiently large, say $|f(s)| \le 2/|s|^2$ if $|s| \le 2|a|$. Moreover,

$$\text{Residue}(0) = \lim_{s\to 0} se^{ts}f(s)$$

$$= \lim_{s\to 0} \frac{e^{ts}}{s-a} = -\frac{1}{a}$$

$$\text{Residue}(a) = \lim_{s\to a} (s-a)e^{ts}f(s)$$

$$= \lim_{s\to a} \frac{e^{ts}}{s} = \frac{e^{at}}{a}$$

Hence

$$f(t) = \text{sum of residues of } e^{ts}f(s) \text{ at all the poles of } f(s)$$

$$f(t) = \frac{1}{a}(e^{at} - 1)$$

Of course, $f(s)$ could have been inverted in this case using partial fractions or a convolution method.

Example 4: Find the residues of $f(s) = \dfrac{1}{s(s^2+a^2)^2}$

Solution: We can use partial fraction method and write:

$$f(s) = \frac{1}{s(s^2+a^2)^2} = \frac{1}{s(s-ai)^2(s+ai)^2}$$

Then $f(s)$ has a simple pole at $s = 0$ and a pole of order 2 at $s = \pm ai$.

$$\text{Residue}(0) = \lim_{s\to 0} se^{ts}f(s)$$

$$= \lim_{s\to 0} \frac{e^{ts}}{(s^2+a^2)^2} = \frac{1}{a^4}$$

$$\text{Residue}(ai) = \lim_{s\to ai} \frac{d}{ds}\left[(s-ai)^2 e^{ts}f(s)\right]$$

$$= \lim_{s\to ai} \frac{d}{ds}\left(\frac{e^{ts}}{s(s+ai)^2}\right) = \frac{it}{4a^3}e^{iat} - \frac{1}{2a^4}e^{iat}$$

Similarly we have

$$\text{Residue}(-ai) = \lim_{s\to -ai}\frac{d}{ds}\Big[(s+ai)^2 e^{ts} f(s)\Big]$$

$$= \lim_{s\to -ai}\frac{d}{ds}\left(\frac{e^{ts}}{s(s-ai)^2}\right) = \frac{-it}{4a^3}e^{-iat} - \frac{1}{2a^4}e^{-iat}$$

$$f(t) = \text{sum of residues of } e^{ts}f(s) \text{ at all the poles of } f(s)$$

$$\text{Residue}(0) + \text{Residue}(ai) + \text{Residue}(-ai)$$

$$= \frac{1}{a^4} + \left(\frac{it}{4a^3} - \frac{1}{2a^4}e^{iat}\right) + \left(\frac{-it}{4a^3}e^{-iat} - \frac{1}{2a^4}e^{-iat}\right)$$

$$= \frac{1}{a^4} + \frac{1}{4a^3}\left(e^{iat} - e^{-iat}\right) - \frac{1}{2a^4}\left(e^{iat} - e^{-iat}\right)$$

$$= \frac{1}{a^4}\left(1 - \frac{a}{2}t\sin at - \cos at\right)$$

$$= f(t)$$

E.32 A Sufficient Condition for the Integral Around Γ

The validity of the result Eq. E.73 hinges on the assumption that the integral around Γ in Eq. E.72 approaches zero as $R \to \infty$. A sufficient condition under which this assumption is correct is supplied in the following [3];

Example 1

Theorem: If we can find constants $M > 0, k > 0$ such that on Γ (where $s = Re^{i\theta}$),

$$|f(s)| < \frac{M}{R^k} \tag{E.74}$$

Then the integral around Γ of $s^{st}f(s)$ approaches zero as $R \to \infty$, i.e.,

$$\lim_{R\to\infty}\int_{\Gamma} e^{st}f(s)ds = 0 \tag{E.75}$$

The condition Equation E-48 always holds if $f(s) = p(s)/q(s)$ where $p(s)$ and $q(s)$ are polynomials and the degree of $p(s)$ is less than the degree of $q(s)$. See example below

Example 1: Let $f(s) = p(s)/q(s)$ where $p(s)$ and $q(s)$ are polynomials such that the degree of $p(s)$ is less than the degree of $q(s)$. Prove that $f(s)$ satisfies the condition of in Example 1 of Sect. 2.7.

Solution: Let

$$p(s) = a_0 s^m + a_1 s^{m-1} + \cdots + a_m$$

$$q(s) = b_0 s^m + b_1 s^{n-1} + \cdots + b_n$$

Where $a \neq 0$, $b \neq 0$ and $0 \leq m \leq n$. Then if $s = Re^{i\theta}$, we have

$$\begin{aligned} f(s) &= \left|\frac{P(s)}{q(s)}\right| = \left|\frac{a_0 s^m + a_1 s^{m-1} + \cdots + a_m}{b_0 s^n + b_1 s^{n-1} + \cdots + b_n}\right| \\ &= \left|\frac{a_0 R^m e^{mi\theta} + a_1 R^{m-1} e^{(m-1)i\theta} + \cdots + a_m}{b_0 R^n e^{ni\theta} + b_1 R^{n-1} e^{(n-1)i\theta} + \cdots + b_n}\right| \\ &= \left|\frac{a_0}{b_0}\right| \frac{1}{R^{n-m}} \left|\frac{1 + \left(\frac{a_1}{a_0 R}\right) e^{-i\theta} + \left(\frac{a_2}{a_0 R^2}\right) e^{-2i\theta} + \cdots + \left(\frac{a_m}{a_0 R^2}\right) e^{-mi\theta}}{1 + \left(\frac{b_1}{b_0 R}\right) e^{-i\theta} + \left(\frac{b_2}{b_0 R^2}\right) e^{-2i\theta} + \cdots + \left(\frac{b_m}{b_0 R^2}\right) e^{-mi\theta}}\right| \end{aligned}$$

Let A denote the maximum of $|a_1/a_0|, |a_2/a_0|, \cdots, |a_m/a_0|$.
Let B denote the maximum of $|b_1/b_0|, |b_2/b_0|, \cdots, |b_m/b_0|$.
Then

$$\begin{aligned} \left|1 + \frac{a_1}{a_0 R} e^{-i\theta} + \frac{a_2}{a_0 R^2} e^{-2i\theta} + \cdots + \frac{a_m}{a_0 R^2} e^{-mi\theta}\right| &\leq 1 + \frac{A}{R} + \frac{A}{R^2} + \cdots + \frac{A}{R^m} \\ &\leq 1 + \frac{A}{R}\left(1 + \frac{1}{R} + \frac{1}{R^2} + \cdots\right) \\ &\leq 1 + \frac{A}{R-1} < 2 \end{aligned}$$

for $R > A + 1$.

Also,

$$\left|1=\frac{b_1}{b_0R}e^{-i\theta}+\frac{b_2}{b_0R^2}e^{-2i\theta}+\cdots+\frac{b_n}{b_0R^2}e^{-ni\theta}\right|\geq 1-\left|\frac{b_1}{b_0R}e^{-i\theta}\frac{b_2}{b_0R^2}e^{-2i\theta}+\cdots+\frac{b_n}{a_0R^2}e^{-ni\theta}\right|$$

$$\geq 1-\left(1+\frac{1}{R}+\frac{1}{R^2}+\cdots+\frac{B}{R^n}\right)$$

$$\geq 1+\frac{B}{R}\left(1+\frac{1}{R}+\frac{1}{R^2}+\cdots\right)$$

$$\geq 1-\frac{B}{R-1}\geq\frac{1}{2}$$

for $R > 2B + 1$

Thus for R larger than either $A + 1$ or $2B + 1$, we have

$$f(s) \leq \left|\frac{a_0}{b_0}\right| \cdot \frac{1}{R^{n-m}} \cdot \frac{1}{1/2} \leq \frac{M}{R^k}$$

Where M is any constant greater than $2|a_0/b_0|$ and $k = n - m \geq 1$. This proves the required results.

E.33 Modification of Bromwich Contour in case of Branch Points

If $f(s)$ has branch points, extensions of the above results can be made provided that the Bromwich contour is suitably modified. For example, if $f(s)$ has only one branch point at $s = 0$, then we can use the contour of Fig. E.5. In this figure, *BDE* and *LNA* represent arcs of a circle of radius R with center at origin O, while *HJK* is the arc of a circle of radius ε with center at O. For details of evaluating inverse Laplace transform in such case see below example for case of Inverse Laplace Transform of function with branch points.

Example 1: Find $\pounds^{-1}\left\{\frac{e^{-a\sqrt{s}}}{s}\right\}$ by use of the complex inversion formula.

Solution: By the complex inversion formula, the required inverse Laplace transform is given by

$$f(t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{st-a\sqrt{s}}}{s} ds \tag{E.76}$$

Since $s = 0$ is a branch point of the integrand, we consider

$$\frac{1}{2\pi i}\oint_c \frac{e^{st-a\sqrt{s}}}{s}ds = \frac{1}{2\pi i}\int_{AB}\frac{e^{st-a\sqrt{s}}}{s}ds + \frac{1}{2\pi i}\int_{BDE}\frac{e^{st-a\sqrt{s}}}{s}ds + \frac{1}{2\pi i}\int_{ER}\frac{e^{st-a\sqrt{s}}}{s}ds$$
$$+ \frac{1}{2\pi i}\int_{HJK}\frac{e^{st-a\sqrt{s}}}{s}ds + \frac{1}{2\pi i}\int_{KL}\frac{e^{st-a\sqrt{s}}}{s}ds + \frac{1}{2\pi i}\int_{LNA}\frac{e^{st-a\sqrt{s}}}{s}ds$$

where C is the contour of Fig. E.5 consisting of the line $AB(s=\gamma)$, the arcs BDE and LNA of a circle of radius R and center at origin O, and the arc HJK of a circle ε with center at O.

Since the only singularity $s=0$ of the integrand is not inside C, the integral on the left is zero by Cauchy's theorem. Also, the integrand satisfies the condition of Example 1 of Sect. 2.7 so that on taking the limit as $R\to\infty$ the integrals along BDE and LNA approach zero. It follows:

$$f(t) = \lim_{\substack{R\to\infty\\ \varepsilon\to 0}} \frac{1}{2\pi i}\int_{AB}\frac{e^{st-a\sqrt{s}}}{s}ds = \frac{1}{2\pi i}\int_{\gamma+i\infty}^{\gamma+i\infty}\frac{e^{st-a\sqrt{s}}}{s}ds$$
$$= \lim_{\substack{R\to\infty\\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{\int_{EH}\frac{e^{st-a\sqrt{s}}}{s}ds + \int_{HJK}\frac{e^{st-a\sqrt{s}}}{s}ds + \int_{KL}\frac{e^{st-a\sqrt{s}}}{s}ds\right\} \tag{E.77}$$

Along EH, $s=xe^{\pi i}$, $\sqrt{s}=\sqrt{xe^{\pi i/2}}=i\sqrt{x}$ and as s goes from $-R$ to $-\varepsilon$, x goes from R to ε. Hence we have

$$\int_{EH}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{-\varepsilon}^{-R}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{R}^{\varepsilon}\frac{e^{-xt-ai\sqrt{x}}}{x}dx$$

Similarly, along KL, $s=xe^{\pi i}$, $\sqrt{s}=\sqrt{xe^{-\pi/2}}=-i\sqrt{x}$ and as s goes from $-\varepsilon$ to $-R$, x goes from ε to R. Then

$$\int_{KL}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{-\varepsilon}^{-R}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{\varepsilon}^{R}\frac{e^{-xt+ai\sqrt{x}}}{x}dx$$

Along HJK, $s=\varepsilon e^{i\theta}$ and we have

$$\int_{HJK}\frac{e^{st-a\sqrt{s}}}{s}ds = \int_{\pi}^{-\pi}\frac{e^{\varepsilon e^{i\theta}t-a\sqrt{\varepsilon}e^{i\theta/2}}}{\varepsilon e^{i\theta}}\varepsilon e^{i\theta}d\theta$$
$$= i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta}t-a\sqrt{\varepsilon}e^{i\theta/2}}d\theta$$

Thus Eq. (E.77) becomes

$$f(t) = -\lim_{\substack{R\to\infty\\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{\int_R^{\varepsilon} \frac{e^{-xt-ai\sqrt{x}}}{x}dx + \int_{\varepsilon}^{R} \frac{e^{-xt+ai\sqrt{x}}}{x}dx + i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta}t - a\sqrt{\varepsilon e^{i\theta}}/2}d\theta\right\}$$

$$= -\lim_{\substack{R\to\infty\\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{\int_{\varepsilon}^{R} \frac{e^{-xt}\left(e^{ai\sqrt{x}} - e^{-ai\sqrt{x}}\right)}{x}dx + i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta}t - a\sqrt{\varepsilon e^{i\theta}}/2}d\theta\right\}$$

$$= -\lim_{\substack{R\to\infty\\ \varepsilon\to 0}} \frac{1}{2\pi i}\left\{2i\int_{\varepsilon}^{R} \frac{e^{-xt}\sin a\sqrt{x}}{x}dx + i\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta}t - a\sqrt{\varepsilon e^{i\theta}}/2}d\theta\right\}$$

Since the limit can be taken underneath the integral sign, we have

$$\lim_{\varepsilon\to 0}\int_{\pi}^{-\pi} e^{\varepsilon e^{i\theta}t - a\sqrt{\varepsilon e^{i\theta}}/2}d\theta = \int_{\pi}^{-\pi} 1 d\theta = -2\pi$$

and so we find

$$f(t) = 1 - \frac{1}{\pi}\int_0^{\infty} \frac{e^{-xt}\sin a\sqrt{x}}{x}dx$$

This can be written as follows (See Example 2 below).

$$f(t) = 1 - erf\left(\frac{a}{2\sqrt{t}}\right)erfc\left(\frac{a}{2\sqrt{t}}\right)$$

Example 2: Prove that $\frac{1}{\pi}\int_0^{\infty} \frac{e^{-xt}\sin a\sqrt{x}}{x}dx = erf\left(\frac{a}{2\sqrt{t}}\right)$ and thus establish the final result in pervious example.

Solution: Letting $x = u^2$, the required integral becomes

$$I = \frac{2}{\pi}\int_0^{\infty} \frac{e^{-u^2t}\sin au}{u}du$$

Then differentiating with respect to a we have

$$\frac{\partial I}{\partial a} = \frac{2}{\pi}\int_0^{\infty} e^{-u^2t}\cos(au)du = \frac{2}{\pi}\left(\frac{\sqrt{\pi}}{2\sqrt{t}}e^{-a^2/4t}\right) = \frac{1}{\sqrt{\pi t}}e^{-a^2/4t}$$

Hence, using the fact that $I = 0$ when $a = 0$,

$$I = \int_0^a \frac{1}{\sqrt{\pi t}} e^{-p^2/4t} dp = \frac{2}{\sqrt{\pi}} \int_0^{a/2\sqrt{t}} e^{-u^2} du = erf\left(a/2\sqrt{t}\right)$$

and the required result is established.

E.34 Case of Infinitely Many Singularities

If we are planning on finding the inverse Laplace transform of functions which have infinitely many isolated singularities, the above methods can be applied [3]. In such case the curved portion of the Bromwich contour is chosen to be of such radius R_m so as to enclose only a required inverse Laplace transform is then found by taking an appropriate limit as $m \to \infty$.

E.35 Inverse Laplace Transform of Functions with Infinitely Many Singularities

We present few examples of a function with infinitely many singularities by first finding its residue then showing the Inverse Laplace Transform of such function as follows:

Example 1: Find all the singularities of $f(s) = \dfrac{\cosh x\sqrt{s}}{s\cosh\sqrt{s}}$ where $0 < x < 1$.

Solution: Because of the presence of $\sqrt{s}$, it would appear that $s = 0$ is a branch point. That this is not so, however, can be seen by noting that

$$f(s) = \frac{\cosh\ x\sqrt{s}}{s\cosh\sqrt{s}} = \frac{1 + (x\sqrt{s})^2 2! + (x\sqrt{s})^4 4! + \cdots}{s\left\{1 + (\sqrt{s})^2 + 2!, +(\sqrt{s})^4, 4, !, +, \cdots\right\}}$$

$$= \frac{1 + x^2 s/2! + x^4 s/4! + \cdots}{s\{1 + x^2 s/2! + s^4/4! + \cdots\}}$$

from which it is evident that there is no branch point as $s = 0$. However, there is a simple pole at $s = 0$. The function $f(s)$ also has infinitely many poles given by the root of the equation

$$\cosh\sqrt{s} = \frac{e^{\sqrt{s}} + e^{-\sqrt{s}}}{2} = 0$$

These occur where $e^{2\sqrt{s}} = -1 = e^{\pi i + 2k\pi i}$ $k = 0, \pm 1, \pm 2, \cdots$

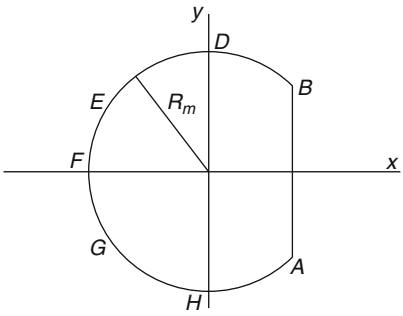


Fig. E.8 Different Functions Configurations

from which $\sqrt{s} = \left(k + \frac{1}{2}\right)\pi i$ or $s = -\left(k + \frac{1}{2}\right)^2 \pi^2$

Thus $f(s)$ has simple poles at

$s = 0$ and $s = s_n$ where $s_n = -\left(\pi - \frac{1}{2}\right)^2 \pi^2 \; n = 1, 2, 3, \cdots$

Example 2: Find $\pounds^{-1}\left\{\dfrac{\cosh x\sqrt{s}}{s\cosh\sqrt{s}}\right\}$ where $0 < x < 1$.

Solution: The required inverse can be bound by using the Bromwich contour of Fig. E.6. The AB is chosen so as to lie to the right of all the poles which, as seen in Example 1 above, are given by (Fig. E.8)

$s = 0$ and $s = s_n$ where $s = -\left(\pi - \frac{1}{2}\right)^2 \pi^2 \; n = 1, 2, 3, \cdots$

We choose the Bromwich contour so that the curved portion $BDEFGHA$ is an arc of a circle Γ_m with center at the origin and radius.

$$R_m = m^2\pi^2$$

where m is a positive integer. This choice insures that the contour does not pass through any of the poles.

We now find the residues of

$$\frac{e^{st}\cosh x\sqrt{s}}{s\cosh\sqrt{s}}$$

at the poles. We have

Residue at $s = 0$ is $\lim\limits_{s\to 0}(s-0)\left\{\dfrac{e^{st}\cosh x\sqrt{s}}{s\cosh\sqrt{s}}\right\} = 1$

Residue at $s_n = -\left(\pi - \frac{1}{2}\right)^2 \pi^2, \quad n = 1, 2, 3, \cdots$ is

$$\lim_{s\to 0}(s-s_n)\left\{\frac{e^{st}\cosh x\sqrt{s}}{s\cosh\sqrt{s}}\right\} = \lim_{s\to 0}\left\{\frac{(s-s_n)}{\cosh\sqrt{s}}\right\}\lim_{s\to 0}\left\{\frac{e^{st}\cosh x\sqrt{s}}{s}\right\}$$

$$= \lim_{s\to 0}\left\{\frac{1}{(\sinh\sqrt{s})(1/2\sqrt{s})}\right\} = \lim_{s\to 0}\left\{\frac{e^{st}\cosh x\sqrt{s}}{s}\right\}$$

$$= \frac{4(-1)^n}{\pi(2n-1)}e^{-(n-1/2)^2\pi^2 t}\cos\left(n-\frac{1}{2}\right)\pi x$$

If C_m is the contour of Fig. E.6, then

$$\frac{1}{2\pi i}\oint_{C_m}\frac{e^{st}\cosh x\sqrt{s}}{s\cosh\sqrt{s}}ds = 1+\frac{4}{\pi}\sum_{n=1}^{m}\frac{(-1)^n}{(2n-1)}e^{-(n-1/2)^2\pi^2 t}\cos\left(n-\frac{1}{2}\right)\pi x$$

Taking the limit as $m\to\infty$ and noting that the integral around Γ_m approaches zero, we find

$$\pounds^{-1}\left\{\frac{\cosh x\sqrt{s}}{s\cosh\sqrt{s}}\right\} = 1+\frac{4}{\pi}\sum_{n=1}^{m}\frac{(-1)^n}{(2n-1)}e^{-(n-1/2)^2\pi^2 t}\cos\left(n-\frac{1}{2}\right)\pi x$$

$$= 1+\frac{4}{\pi}\sum_{n=1}^{m}\frac{(-1)^n}{(2n-1)}e^{-(2n-1)^2\pi^2 t/4}\cos\left(\frac{(2n-1)\pi x}{2}\right)$$

Example 3: Find $\pounds^{-1}\left\{\dfrac{\sinh sx}{s^2\cosh sa}\right\}$ where $0<x<a$.

Solution: The function $f(s)=\dfrac{\sinh sx}{s^2\cosh sa}$ has poles at $s=0$ and at values of s for , i.e.,

$$s=s_k=-\left(\pi-\frac{1}{2}\right)\pi i/a \quad k=0,\pm 1,\pm 2,\cdots$$

Because of the presence of s^2, it would appear that $s=0$ is a pole of order two. However, by observing that near $s=0$,

$$f(s)=\frac{\sinh sx}{s^2\cosh sa} = \frac{sx+(sx)^3/3!+(sx)^5/5!+\cdots}{s\left\{1+(sa)^2/2!+(sa)^4/4!+\cdots\right\}}$$

$$= \frac{x+s^2x^3/3!+s^4x^5/5!}{s^2\{1+s^2a^2/2!+s^4a^4/4!+\cdots\}}$$

We see that $s = 0$ is a pole of order one, i.e., a simple pole. The poles s_k are also simple poles.

Proceeding at $s = 0$ is as in Example 2 above, we obtain the residues of $e^{st}f(s)$ at these poles.

Residue at $s = 0$ is

$$\lim_{s\to 0}(s-0)\left\{\frac{e^{st}\sinh sx}{s^2\cosh sa}\right\} = \left\{\lim_{s\to 0}\frac{\sinh sx}{s}\right\}\left\{\lim_{s\to 0}\frac{e^{st}}{s^2\cosh sa}\right\} = x$$

Using L'Hospital rule.

Residues at $s = s_k$ is

$$\begin{aligned}\lim_{s\to 0}(s-s_k)\left\{\frac{e^{st}\sinh sx}{s^2\cosh sa}\right\} &= \left\{\lim_{s\to s_k}\frac{s-s_k}{\cos sa}\right\}\left\{\lim_{s\to s_k}\frac{e^{st}\sinh sx}{s^2}\right\}\\ &= \left\{\lim_{s\to s_k}\frac{1}{a\sinh sa}\right\}\left\{\lim_{s\to s_k}\frac{e^{st}\sinh sx}{s^2}\right\}\\ &= \frac{1}{ai\sin(k+1/2)\pi}\cdot\frac{e^{(k+1/2)\pi it/a}i\sin(k+1/2)(\pi x/a)}{-(k+1/2)^2\pi^2/a^2}\\ &= -\frac{a(-1)^k e^{(k+1/2)\pi it/a}i\sin(k+1/2)(\pi x/a)}{\pi^2(k+1/2)^2}\end{aligned}$$

By and appropriate limiting procedure similar to that used in Example 2 above, we find on taking the some of the residues the required result,

$$\begin{aligned}\pounds^{-1}\left\{\frac{\sinh sx}{s^2\cosh sa}\right\} &= x - \frac{a}{\pi^2}\sum_{k=-\infty}^{\infty}\frac{(-1)^k e^{(k+1/2)\pi it/a}i\sin(k+1/2)(\pi x/a)}{(k+1/2)^2}\\ &= x - \frac{2a}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^k\cos(n-1/2)(\pi t/a)\sin(k+1/2)(\pi x/a)}{(k+1/2)^2}\\ &= x + \frac{8a}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^k}{(k+1/2)^2}\sin\frac{(2n-1)\pi x}{2a}\cos\frac{(2n-1)\pi t}{2a}\end{aligned}$$

E.36 Inverse Laplace Transform of Functions with Branch Point

See Sect. 2.10 of this appendix and related examples that are presented there. More examples can be found in [4] and [7] of this appendix. We recommend the reader to refer themselves to them

Example 1: Let $f(s) = \dfrac{p(s)}{q(s)}$ where $p(s)$ and $q(s)$ are polynomials (having no common roots) of degree n and m, respectively, $m > n$, and $q(s)$ has simple roots at $z_1, z_2, \cdots z_m$. Then $f(s)$ has a simple pole at each $s = z_k$, and writing

$$\begin{aligned} f(s) &= \frac{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0}{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_0} \\ &= \frac{a_n + \dfrac{a_{n-1}}{s} + \cdots + \dfrac{a_0}{s}}{s^{m-n}\left(b_m + \dfrac{b_{m-1}}{s} + \cdots + \dfrac{b_0}{s^m}\right)} \qquad (a_n, b_m \neq 0) \end{aligned}$$

Solution: It is enough to observe that for $|s|$ suitably large

$$\left|a_n + \frac{a_{n-1}}{s} + \cdots + \frac{a_0}{s^n}\right| \leq |a_n| + |a_{n-1}| + \cdots + |a_0| = c_1$$

$$\left|b_m + \frac{b_{m-1}}{s} + \cdots + \frac{b_0}{s^m}\right| \geq |b_m| - \frac{|b_{m-1}|}{|s|} - \cdots - \frac{|b_0|}{|s^m|} \geq \frac{|b_m|}{2} = c_2$$

and thus

$$|f(s)| \leq \frac{c_1/c_2}{|s|^{m-n}}$$

Hence by Example 1 above, we have

$$\text{Res}(z_k) = \frac{e^{z_k t}p(z_k)}{q'(z_k)} \quad k = 1, 2, \cdots, m$$

and

$$f(t) = \sum_{k=1}^{m} \frac{p(z_k)}{q'(z_k)} e^{z_k t}$$

E.37 References

1. Fourier Transformation - Wikipedia, the free encyclopedia
2. Spiegel MR. Fourier analysis with applications to boundary value problem. Schaum's Outline, McGraw Hill
3. Spiegel MR. Laplace transforms. Schaum's Outline, McGraw Hill
4. Schiff JL (1991) The laplace transform, theory and applications. Springer
5. Gerald K (1994) A Friendly Guide to Wavelets. Birkhäuser, ISBN 0-8176-3711-7
6. Elias S, Rami S (2003) Fourier analysis: an introduction. Princeton University Press, ISBN 0-691-11384-X
7. Dennis Z. (2001) A first course in differential equations with modeling applications. 7th Edition, Brooks/Cole Thomson Learning

Appendix F
Short Course in Electromagnetic

Electromagnetism is one of the four fundamental interactions of nature, along with strong interaction, weak interaction, and gravitation. It is the force that causes the interaction between electrically charged particles; the areas in which this happens are called electromagnetic fields (Fig. F.1).

Electromagnetism is the force responsible for practically all the phenomena encountered in daily life (with the exception of gravity). Ordinary matter takes its form as a result of intermolecular forces between individual molecules in matter. Electromagnetism is also the force which holds electrons and protons together inside atoms, which are the building blocks of molecules. This governs the processes involved in chemistry, which arise from interactions between the electrons orbiting atoms.

Electromagnetic field theory has been and will continue to be one of the most important fundamental courses of electrical engineering curriculum. It is one of the best-established general theories that provides explanations and solutions to intricate electrical engineering problems when other theories are no longer applicable

F.1 Maxwell's Equations and Electric Field of the Electromagnetic Wave

Although Maxwell's equations were formulated over one hundred years ago, the subject of electromagnetism has not remained static. Much of modern physics (and engineering) involves time-dependent electromagnetic fields in which Maxwell's displacement current plays a crucial role. Maxwell's equations contain all the information necessary to characterize the electromagnetic fields at any point in a medium.

In order to understand behavior of materials response to a high power laser beam one needs to consider the coupling of the laser energy to a material. Therefore, we

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

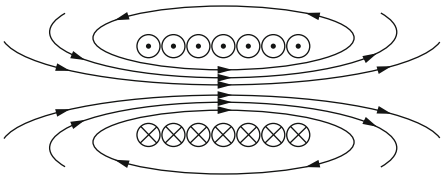


Fig. F.1 Electricity-Magnetism

need first to know the optical reflectivity R and the transmissivity T for light incident on a surface which divides two semi-infinite media. To understand reflectivity, we must use some general results from the theory of electromagnetic waves.

For the **E**lectro **M**agnetic (EM) fields to exist they must satisfy the four Maxwell equations at the source where they are generated, at any point in a medium through which they propagate, and at the load where they received or absorbed As the fields must satisfy the four coupled Maxwell equations involving four unknown variables, we first obtain an equation in terms of one unknown variables. Similar equations can then be obtained for the other variables. We refer to these equations as the general *Wave Equations*. It can be shown that the fields generated by time-varying sources propagate as *Spherical Waves*. However, in a small region far away from the radiating source, the spherical wave may be approximated as a *Plane Wave*, that is, one in which all the field quantities are in a plane normal to the direction of its propagation (the transverse plane). Consequently, a plane wave does not have any field component in its direction of propagation (the longitudinal direction).

We first seek the solution of a plane wave in an unbounded dielectric medium and show that the wave travels with the speed of light in free space. We then consider the general case of a finitely conducting medium. We show that the wave attenuates as a result of loss in energy as it travels in the conducting medium. Finally, we introduce the concept of reflection and transmission of a plane wave when it leaves one medium and enters another.

F.2 The Wave Equations for Electric and Magnetic Field

In regions of space where there is no charge or current, Maxwell's equations read

$$\left.\begin{aligned} &\text{(i)}\ \ \nabla \cdot \vec{E} = 0 \quad \text{(iii)} \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \\ &\text{(ii)} \nabla \cdot \vec{B} = 0 \quad \text{(iv)} \nabla \times \vec{B} = \mu_0 \varepsilon_0 - \frac{\partial \vec{E}}{\partial t} \end{aligned}\right\} \tag{F.1}$$

where

E is the Electric Field

B is the Magnetic Field

μ_0 is the constant Biot–Savart's law known as permittivity of free space $4\pi \times 10^7\,\mathrm{N/A^2}$

ε_0 is the constant Coulomb's law known as permittivity of free space $8.85 \times 10^{-12}\,\mathrm{C^2/(N \cdot m^2)}$

The above equations are constitute a set of coupled, first-order, partial differential equations for $\vec{E}$ and $\vec{B}$. They can be decoupled by applying the curl to (iii) and (iv):

$$\nabla \times \left(\nabla \times \vec{E}\right) = \nabla\left(\nabla \cdot \vec{E}\right) - \nabla \times \left(-\frac{\partial \vec{B}}{\partial t}\right)$$

$$= -\frac{\partial}{\partial t}\left(\nabla \times \vec{B}\right) = -\mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

$$\nabla \times \left(\nabla \times \vec{B}\right) = \nabla\left(\nabla \cdot \vec{B}\right) - \nabla \times \left(\mu_0 \varepsilon_0 - \frac{\partial \vec{E}}{\partial t}\right)$$

$$= \mu_0 \varepsilon_0 \frac{\partial}{\partial t}\left(\nabla \times \vec{E}\right) = -\mu_0 \varepsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}$$

Or, since $\nabla \cdot \vec{E} = 0$ and $\nabla \cdot \vec{B} = 0$.

$$\nabla^2 \vec{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} \quad \nabla^2 \vec{B} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2} \tag{F.2}$$

Equation F.2 is a demonstration of separations between $\vec{E}$ and $\vec{B}$, but they are second order [1]. In vacuum, then, each Cartesian component of $\vec{E}$ and $\vec{B}$ satisfies the **Three-Dimensional Wave Eq.** F.3

$$\nabla^2 f = \frac{1}{v^2} \frac{\partial^2 f}{\partial t^2} \tag{F.3}$$

This supports standard wave equation in general within Cartesian form, which is similar to a classical wave equation of small disturbance on the string, where v represents the speed of propagation and is given by [1];

$$v = \sqrt{\frac{T}{\mu}} \tag{F.4}$$

Where μ the mass per unit length and this equation is admits as solutions all functions of the form as Eq. F.5 for example in z direction of propagation [1].

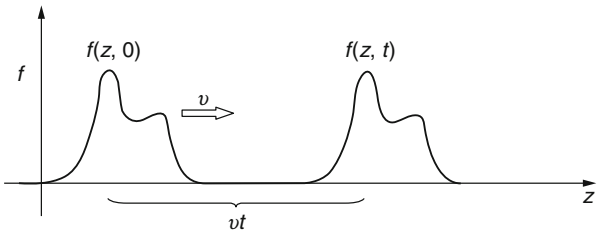


Fig. F.2 Wave Propagation in z-direction [1]

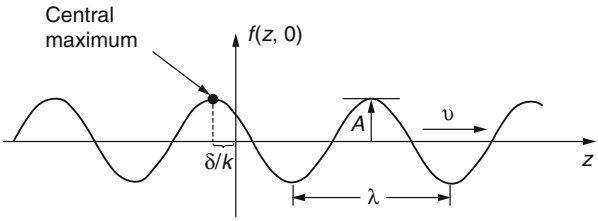


Fig. F.3 A Sinusoidal wave propagation [1]

$$f(z,t) = g(z - \upsilon t) \tag{F.5}$$

This mathematical derivation can be sort of depicted as Fig. F.2 at two different two times, once at $t = 0$, and at later time t. Each point on the wave form simply shifts to the right by an amount υt, where υ is the velocity.

Then we can see that Maxwell's equations imply that empty space supports the propagation of electromagnetic waves, traveling at a speed

$$\upsilon = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 3.00 \times 10^8 \, meter/\sec ond = 3.00 \times 10^8 m/s = c$$

which is precisely the velocity of light c [1], which Maxwell himself was astonished and he, himself put it "wave can scarcely avoid the interference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena".

F.3 Sinusoidal Waves

Of all possible wave forms, the sinusoidal one form like Eq. F.6 below is for a good reason the most familiar. Figure F.3 shows this function at time $t = 0$ with A as **amplitude** of the wave.

$$f(z,t) = A\cos\left[k(z - \upsilon t) + \delta\right] \tag{F.6}$$

(it is positive, and represents the maximum displacement from equilibrium). The argument of the cosine is called the **phase**, and δ is the **phase constant** and one can add any integer multiple 2π to δ without changing $f(z,t)$. Ordinarily, one uses a value in the range $0 \leq \delta < 2\pi$. A point known as Central Maximum on the curve of takes place at $z = \upsilon t - \delta/k$ when the phase is zero. If $\delta = 0$, the central maximum passes the origin at time $t = 0$; more generally, δ/k is the distance by which the central maximum (and therefore the entire wave) is "delayed". Finally, k is the **wave number**; It is related to the **wavelength** λ by the Eq. F.7 for when z advances by $2\pi/k$, and the cosine executes one complete cycle.

$$\lambda = \frac{2\pi}{k} \tag{F.7}$$

As time passes, the entire wave train proceeds to the right, at speed υ. At any fixed point z, the string vibrates up and down, undergoing one full cycle in a **period**.

$$T = \frac{2\pi}{k\upsilon} \tag{F.8}$$

If now introduce frequency v as number of oscillations per unit time and show it in the form of Eq. F.9, then we have:

$$v = \frac{1}{T} = \frac{k\upsilon}{2\pi} = \frac{\upsilon}{\lambda} \tag{F.9}$$

For our purpose it is nicer to write the Eq. F.6 in a nicer form and present that in terms of **angular frequency** ω given below in the analogous case of uniform circular motion that represents the number of radians swept out per unit time.

$$\omega = 2\pi v = k\upsilon \tag{F.10}$$

Then the new form of sinusoidal wave in terms of ω rather than υ is given as below:

$$f(z,t) = A\cos(kz - \omega t + \delta) \tag{F.11}$$

If both k and ω traveling to the left then Eq. F.11 can be written in the form of [1];

$$f(z,t) = A\cos(kz + \omega t - \delta) \tag{F.12}$$

This in correspondence to and consistent with our previous convention that δ/k shall represent the distance by which the wave is "delayed" since the wave is now moving to the *left*, and delay means a shift to the *right*. At $t = 0$, the wave looks like Fig. F.4 and since the cosine is an *even function*, we can just as well write Eq. F.12 as form of Eq. F.13 as below.

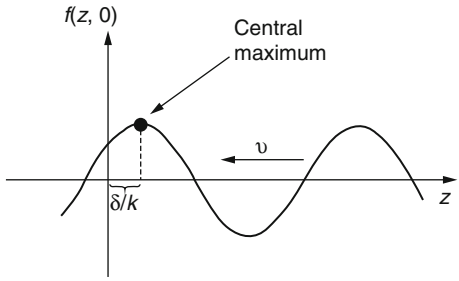


Fig. F.4 Sinusoidal wave traveling to the left [1]

$$f(z,t) = A\cos(-kz - \omega t + \delta) \tag{F.13}$$

Comparing both Eqs. of 12 and 13 we conclude that *we could simply switch the sign of k* to produce a wave with the same amplitude, phase constant, frequency, and wavelength traveling in opposite direction.

What we have learned in Appendix D about complex in view of Euler's formula, we have

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{F.14}$$

We can now write Eq. C.16 like below;

$$f(z,t) = \mathrm{Re}\left[Ae^{i(kz-\omega t+\delta)}\right] \tag{F.15}$$

where Re[ξ] denotes the real part of any complex number such as ξ. This allows us to introduce the **Complex Wave Function** of Eq. F.16 as below;

$$\widetilde{f}(z,t) = \widetilde{A}e^{i(kz-\omega t)} \tag{F.16}$$

with introduction to complex amplitude $\widetilde{A} = Ae^{i\delta}$ absorbing the phase constant. The actual wave function is the real part of $\widetilde{f}$:

$$f(z,t) = \mathrm{Re}\left[\widetilde{f}(z,t)\right] \tag{F.17}$$

Knowing $\widetilde{f}$, it is a simple matter to find f.

Example 1: Combine two sinusoidal waves of f_1 and f_2.

Solution: Let us write the following function f_3 as

$$f_3 = f_1 + f_2 = \mathrm{Re}\left(\widetilde{f}_1\right) + \mathrm{Re}\left(\widetilde{f}_2\right) = \mathrm{Re}\left(\widetilde{f}_1 + \widetilde{f}_2\right) = \mathrm{Re}\left(\widetilde{f}_3\right)$$

with $\widetilde{f}_3 = \widetilde{f}_1 + \widetilde{f}_2$. You may simply add the corresponding complex wave functions, and then take the real part. In particular, if they have the same frequency and wave number,

$$\widetilde{f}_3 = \widetilde{A}_1 e^{i(kz-\omega t)} + \widetilde{A}_2 e^{i(kz-\omega t)} = \widetilde{A}_3 e^{i(kz-\omega t)}$$

where

$$\widetilde{A}_3 = \widetilde{A}_1 + \widetilde{A}_2 \quad \text{or} \qquad \widetilde{A}_3 e^{i\delta_3} = \widetilde{A}_1 e^{i\delta_1} + \widetilde{A}_2 e^{i\delta l 2}$$

Now we are going to figure out what A_3 and δ_3 are as follows:

$$(A_3)^2 = (A_3 e^{i\delta_3})(A_3 e^{i\delta_3}) = (A_1 e^{i\delta_1} + A_2 e^{i\delta_2})(A_1 e^{i\delta_1} + A_2 e^{i\delta_2})$$

$$= (A_1)^2 + (A_2)^2 + A_1 A_2 (e^{i\delta_1} e^{i\delta_2} + e^{i\delta_1} e^{i\delta_2}) = (A_1)^2 + (A_2)^2 + A_1 A_2 \cos(\delta_1 - \delta_2)$$

$$A_3 = \sqrt{(A_1)^2 + (A_2)^2 + A_1 A_2 \cos(\delta_1 - \delta_2)}$$

$$A_3 e^{i\delta_3} = A_3(\cos\delta_3 + i\sin\delta_3) A_1(\cos\delta_1 + i\sin\delta_1) + A_2(\cos\delta_2 + i\sin\delta_2)$$

$$= (A_1\cos\delta_1 + A_2\cos\delta_2) + i(A_1\sin\delta_1 + A_2\sin\delta_2)$$

$$\tan\delta_3 = \frac{A_3\sin\delta_3}{A_3\cos\delta_3} = \frac{A_1\sin\delta_1 + A_2\sin\delta_2}{A_1\cos\delta_1 + A_2\cos\delta_2}$$

$$\delta_3 = \tan^{-1}\left(\frac{A_1\sin\delta_1 + A_2\sin\delta_2}{A_1\cos\delta_1 + A_2\cos\delta_2}\right)$$

As we can see the combined wave still has the same frequency and wavelength is given by

$$f_3(z,t) = A_3\cos(kz - \omega t + \delta_3)$$

F.4 Polarization of the Wave

Polarization, also called wave polarization, is an expression of the orientation of the lines of electric flux in an electromagnetic field (EM field). Polarization can be constant—that is, existing in a particular orientation at all times, or it can rotate with each wave cycle.

Polarization affects the propagation of EM fields at infrared (IR), visible, ultraviolet (UV), and even X-ray wavelengths. In ordinary visible light, there are numerous wave components at random polarization angles. When such light is passed through a special filter, the filter blocks all light except that having a certain

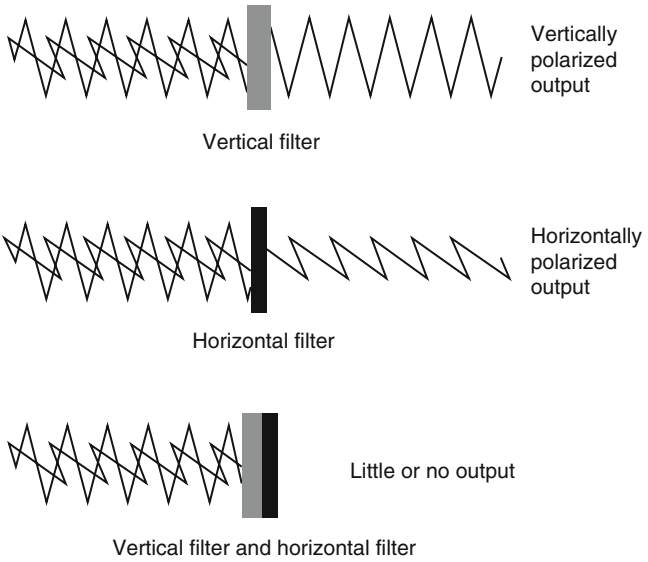


Fig. F.5 Polarization wave

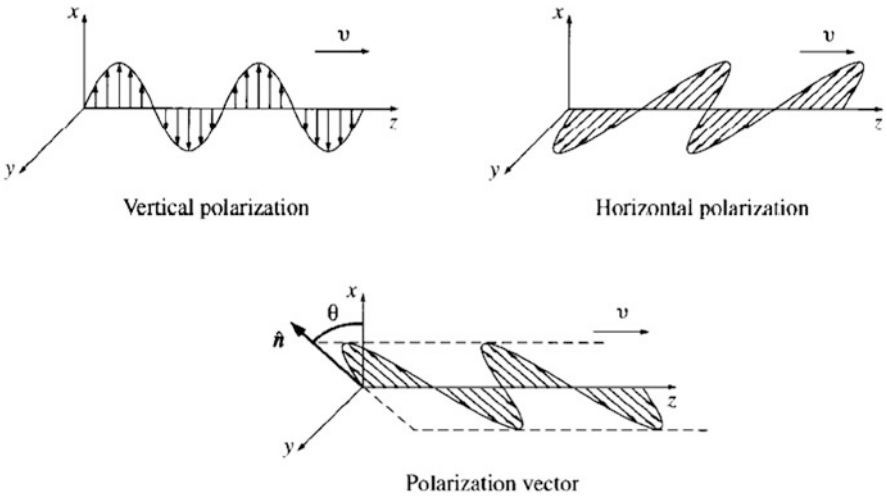


Fig. F.6 Vector form of polarization wave [1]

polarization. When two polarizing filters are placed so that a ray of light passes through them both, the amount of light transmitted depends on the angle of the polarizing filters with respect to each other. The most light is transmitted when the two filters are oriented so they polarize light in the same direction. The least light is transmitted when the filters are oriented at right angles to each other (Fig. F.5) and in vector presentation Fig. F.6.

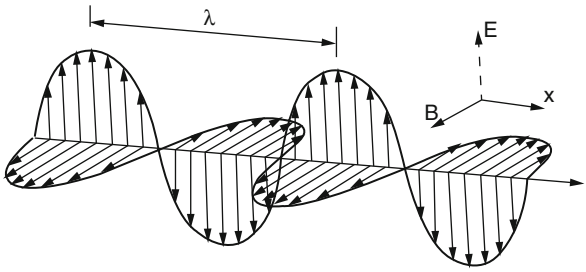


Fig. F.7 Electromagnetic wave propagation

Table F.1 Type of radiation vs. wavelength range

Type of radiation	Wavelength range
Gamma rays	<1 pm
X-rays	1 nm–1 pm
Ultraviolet	400 nm–1 nm
Visible	750 nm–400 nm
Infrared	2.5 mm–750 nm
Microwaves	1 mm–25 mm
Radio waves	>1 mm

Electromagnetic waves are considered transverse waves in which the vibrations of electric and magnetic fields are perpendicular to each other and to the direction of propagation (Fig. F.7). These two fields change with time and space in a sinusoid fashion. Generally, only the electric field is represented, related to the propagation direction, because it is with the electric field that detectors (eye, photographic film, CCD, etc.) interact.

Visible light makes up just a small part of the full electromagnetic spectrum. Electromagnetic waves with shorter wavelengths and higher energies include ultraviolet light, x-rays, and gamma rays. Electromagnetic waves with longer wavelengths and lower energies include infrared light, microwaves, and radio and television waves. Table F.1 below is presentation of type of radiation along with their wave length range.

The polarization of an electromagnetic wave refers to the orientation of its electric field $\vec{E}$. When the direction of electric field ($\vec{E}$) is randomly varying with time on a very fast scale, related to the direction of propagation the wave is considered non-polarized.

In case of a linearly polarized wave, the electric vector has a fixed orientation related to the propagation direction like in Fig. F.8.

The polarization of electromagnetic wave can be produced by absorption, scattering, reflection and birefringence.

A linear polarizer is a device that only allows electric field components parallel to a certain direction (called the polarization axis) to pass through. Any electromagnetic wave that comes through such a polarizer is polarized in the direction of the polarization axis. After leaving the polarizer, the wave's polarization ($\vec{E}$-field direction) does not change, and the wave is considered to be linearly polarized.

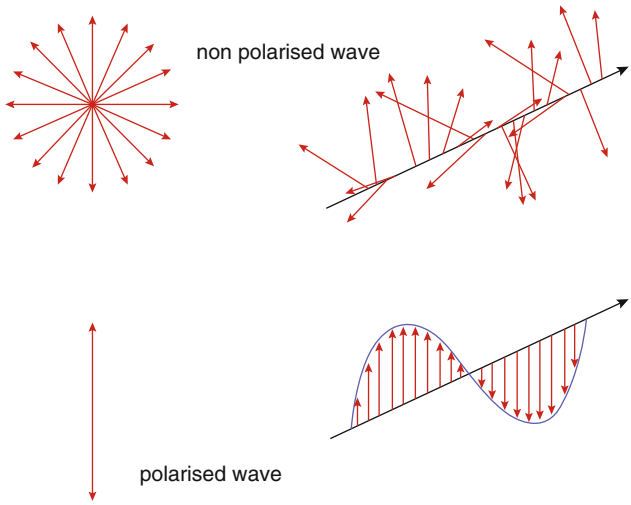


Fig. F.8 Electric field orientation for polarized and non-polarized electromagnetic wave

If the linearly polarized light passes through a second polarizer, the transmitted intensity $I(\theta)$ of the waves as it leaves the second polarizer is given by Malus law.

The specific manner in which a beam of electromagnetic wave at a specific polarization is reflected (and refracted) at interface between two different media can be used to determine the refractive index of the solid.

Specifically, for a particular interface, there is a particular angle of incidence (relative to the normal vector of the surface), called the Brewster angle, which is related to the refractive index of a material. At this angle, the reflection coefficient of light polarized parallel to the plane of incidence is zero. Thus, if the incident light is non-polarized and impinges on the material at the Brewster angle, the light reflected from the solid will be polarized in the plane perpendicular to the plane of incidence. If the incident light is polarized parallel to the plane of incidence, the intensity of the reflected light will be theoretically zero at the Brewster angle.

The proposed experiment uses the polarization by reflection due to its simplicity, but other polarization methods can be used too.

When a light fascicle falls on the M mirror at Brewster angle, the reflected fascicle is linearly polarized. Using a second rotating mirror M' the Malus law can be checked. If mirror M' is rotated around the PP' axes, the reflected P'S' fascicle has a variable intensity with two minim and two maxim values. When the second fascicle falls as the first fascicle at the Brewster angle the S'P' fascicle has a minimum value (See Fig. F.9).

The experiments with visible light and mirrors were made more than one century ago, starting with Brewster and repeated in a lot of laboratories.

To date some experiments were made in order to check the phenomena for IR, and the effect can be accepted as valid for UV and gamma rays. Related to this kind of experiment, it is worth to be reminded a paper from 1969: Rotatable polarizers

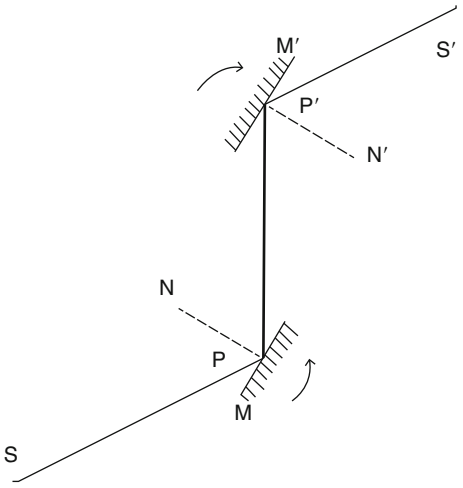


Fig. F.9 Polarization by reflection

for the near infra-red, R M Lambert et al, J. Phys. E: Sci. Instrum. 2 799-801 doi: 10.1088/0022-3735/2/9/311.

Polarization is an important optical property inherent in all laser beams. Brewster windows, reflective phase retarders, and absorbing thin-film reflectors use the advantage of polarization. On the other hand, it can cause troublesome and sometimes unpredictable results when ignored. Since virtually all laser sources exhibit some degree of polarization, understanding this effect is necessary in order to specify components properly. The following text gives a basic polarization definition and presents the polarization types most commonly encountered.

Light is a transverse electromagnetic wave; this means that the electric and magnetic field vectors point perpendicular to the direction of wave travel (Fig. F.10). When all the electric field vectors for a given wave-train lie in a plane, the wave is said to be plane or linearly polarized. The orientation of this plane is the direction of polarization.

Unpolarized light refers to a wave collection which has an equal distribution of electric field orientations for all directions (Fig.F.11). While each individual wave-train may be linearly polarized, there's no preferred direction of polarization when all the waves are averaged together.

Randomly polarized light is exactly what it says; the light is plane polarized, but the direction is unknown, and may vary with time. Random polarization causes problems in optical systems since some components are polarization sensitive. If the polarization state changes with time, then the components' transmission, reflection, and/or absorption characteristics will also vary with time.

Polarization is a vector that has both direction and amplitude. Like any vector, it's defined in an arbitrary coordinate system as the sum of orthogonal components. In Fig. F.12, we see a plane polarized wave which points at 45° to the axes of our coordinate system. Thus, when described in this coordinate system, it has equal x and y components. If we then introduce a phase difference of 90° (or one-quarter

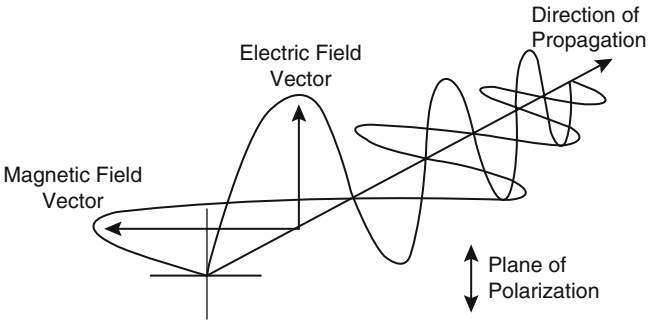


Fig. F.10 Definition of a polarization vector

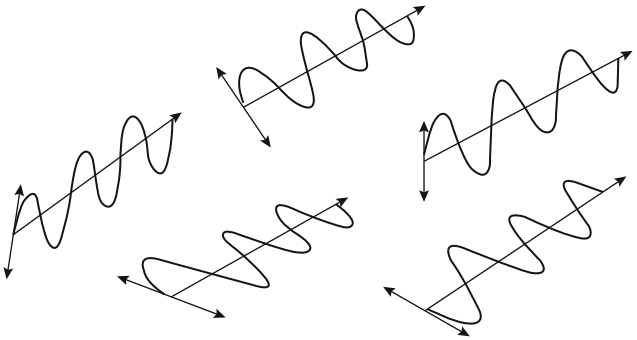
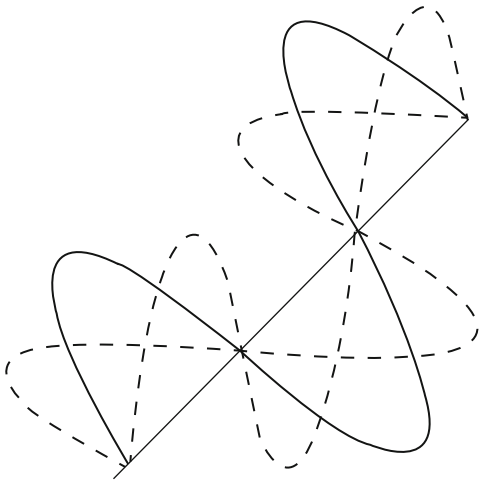


Fig. F.11 Unpolarized light

Fig. F.12 Plane polarized wave which points at 45° to the axes



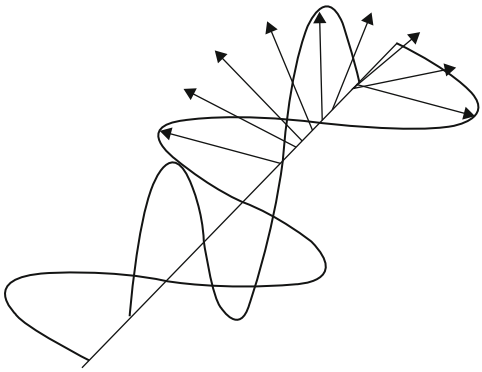


Fig. F.13 Circularly polarized wave

wavelength) between these components, the result is a wave in which the electric field vector has a fixed amplitude but whose direction varies as we move down the wave-train (See Fig. F.13). Such a wave is said to be circularly polarized since the tip of the polarization vector traces out a circle as it passes a fixed point.

If we have two wave-trains with unequal amplitude and with a quarter-wave phase difference, then the result is elliptical polarization. The tip of the polarization vector will trace out an ellipse as the wave passes a fixed point. The ratio of the major to the minor axis is called the ellipticity ratio of the polarization.

Always state the polarization orientation when ordering optical coatings for use at non-normal incidence. If you are unsure about how to determine the polarization state of your source, please contact our applications engineers for assistance.

When light strikes an optical surface, such as a beam splitter, at a non-perpendicular angle, the reflection and transmission characteristics depend upon polarization. In this case, the coordinate system we use is defined by the plane containing the input and reflected beams. Light with a polarization vector lying in this plane is called p-polarized, and light, which is polarized perpendicular to this plane, is called s-polarized. Any arbitrary state of input polarization can be expressed as a vector sum of these s and p components. To understand the significance of s and p polarizations, examine the graph which shows the single surface reflectance as a function of angle of incidence for the s and p components of light at a wavelength of 10.6 μm striking a ZnSe surface (Fig. F.14). Note that while the reflectance of the s component steadily increases with angle, the p component at first decreases to zero at 67° and then increases after that. The angle at which the p reflectance drops to zero is called Brewster's Angle. This effect is exploited in several ways to produce polarizing components or uncoated windows which have no transmission loss such as the Brewster windows.

The angle at which p reflectance drops to zero, termed Brewster's Angle and can be calculated from $\theta_B = \tan^{-1}(n)$, where θ_B is Brewster's Angle and n is the material's index of refraction Polarization state is particularly important in laser cutting application as well as our consideration for purpose of this book and laser interaction with metal. Figure F.15 shows polarization summary.

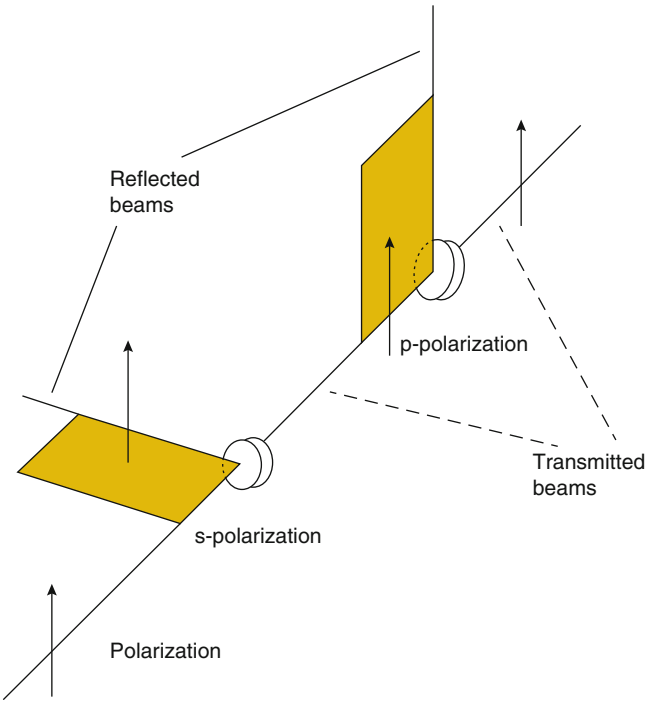


Fig. F.14 For s-polarization, the input polarization is perpendicular to the plane (shown in color) containing the input and output beams. For p-polarization, the input polarization is parallel to the plane (shown in color) containing the input and output beams

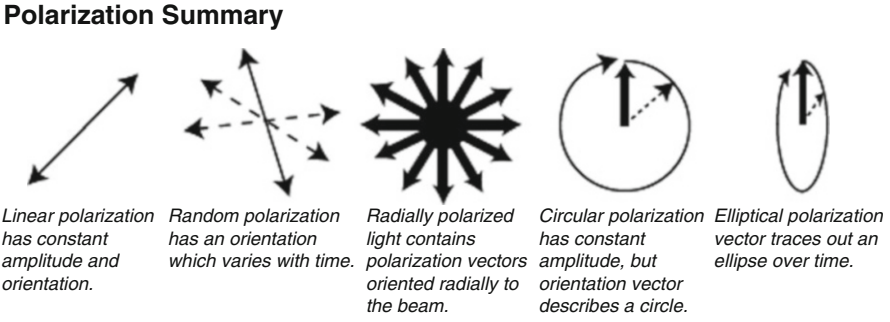

Polarization Summary
Linear polarization has constant amplitude and orientation.
Random polarization has an orientation which varies with time.
Radially polarized light contains polarization vectors oriented radially to the beam.
Circular polarization has constant amplitude, but orientation vector describes a circle.
Elliptical polarization vector traces out an ellipse over time.

Fig. F.15 Polarization Summary

Also by now we know that if the waves travel down the path of let say z-direction is called **Transverse** because the displacement (in case of rope or string) is perpendicular to the direction of propagation and it will be called **Longitudinal** if displacement from equilibrium is along the direction of propagation. Sound waves, which are nothing but compression wave in air, are longitudinal while electromagnetic waves as we shall see are transverse (Section 1.4 of this appendix).

F.5 Monochromatic Plane Waves

As we discussed in Section 1.2 of this appendix we now pay our attention to sinusoidal waves of frequency ω where in visible range, each different frequency corresponds to different *colors*. Such waves are called **Monochromatic Waves** (See Table F.2). Suppose furthermore, that the waves are traveling in the z-direction and have no x or y dependencies; these are called **Plane Waves** [3], because the fields are uniform over every plane perpendicular to the direction of propagation

Table F.2 The electromagnetic spectrum

The electromagnetic spectrum		
Frequency (Hz)	Type	Wavelength (m)
10^{22}		10^{-13}
10^{21}	Gamma rays	10^{-12}
10^{20}		10^{-11}
10^{19}		10^{-10}
10^{18}	X rays	10^{-9}
10^{17}		$\mathbf{10^{-8}}$
10^{17}	Ultraviolet	10^{-7}
10^{15}	Visible	10^{-6}
10^{14}	Infrared	10^{-5}
10^{13}		10^{-4}
10^{12}		10^{-3}
10^{11}		10^{-2}
10^{10}	Microwave	10^{-1}
10^{9}		1
10^{8}	TV, FM	10
10^{7}		10^{2}
10^{6}	AM	10^{3}
10^{5}		10^{4}
10^{4}	RF	10^{5}
10^{3}		10^{6}
The Visible Range		
Frequency **(Hz)**	Color	Wavelength (m)
1.0×10^{15}	Near ultraviolet	3.0×10^{-7}
7.5×10^{14}	Shortest visible blue	$\mathbf{4.0} \times 10^{-7}$
6.5×10^{14}	Blue	$4\,6 \times 10^{-7}$
5.6×10^{14}	Green	5.4×10^{-7}
5.1×10^{14}	Yellow	5.9×10^{-7}
4.9×10^{14}	Orange	6.1×10^{-7}
3.9×10^{14}	Longest visible red	7.6×10^{-7}
3.0×10^{14}	Near infrared	1.0×10^{-6}

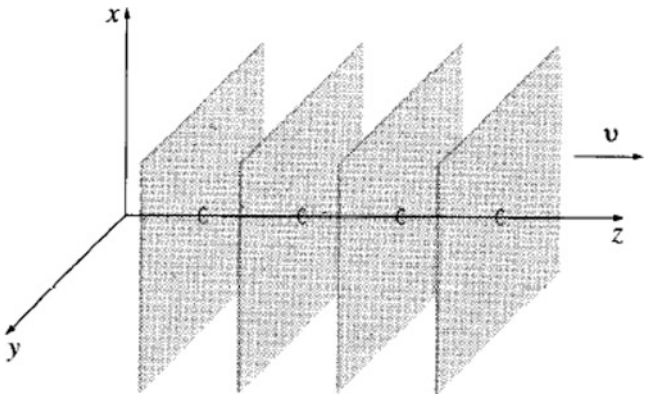


Fig. F.16 Plane wave propagation [1]

(See Fig. F.16). We are interested, and then in complex form of both magnetic and electric filed as follows:

$$\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{i(kx-\omega t)} \quad \text{and} \quad \widetilde{\vec{B}}(z,t) = \widetilde{\vec{B}}_0 e^{i(kx-\omega t)} \tag{F.18}$$

where $\widetilde{\vec{E}}_0$ and $\widetilde{\vec{B}}_0$ are the complex amplitudes [1]. The physical fields, of course are the real parts of $\widetilde{\vec{E}}$ and $\widetilde{\vec{B}}$. Now, the wave equations for $\vec{E}$ and $\vec{B}$ Eq. F.2 were derived from Maxwell's Equations. However, whereas every solution to Maxwell's equations (in empty space) must obey the wave equation, the converse is *not* true; Maxwell's equations impose extra constraints on $\widetilde{\vec{E}}_0$ and $\widetilde{\vec{B}}_0$. In particular, since $\nabla \cdot \vec{E} = 0$ and $\nabla \cdot \vec{B} = 0$, it follows that:

$$\left(\widetilde{E}_0\right)_z = \left(\widetilde{B}_0\right)_z = 0 \tag{F.19}$$

This is because the real part of $\widetilde{\vec{E}}$ differs from the imaginary part only in the replacement of Sine by Cosine, if the former obeys Maxwell's equations, so does the latter, and hence $\widetilde{\vec{E}}$ as well [1]. Equation F.19 also indicates that, *electromagnetic waves are transverse*: the electric and magnetic fields are perpendicular to the direction of propagation. Moreover, Faraday's law, $\nabla \times \vec{E} = -\partial \vec{B}/\partial t$, implies a relationship between the electric and magnetic amplitudes in the following equation;

$$-k\left(\widetilde{E}_0\right)_y = \omega\left(\widetilde{B}_0\right)_x \quad \text{and} \quad -k\left(\widetilde{E}_0\right)_x = \omega\left(\widetilde{B}_0\right)_y \tag{F.20}$$

or in compact form complex variable presentation we have:

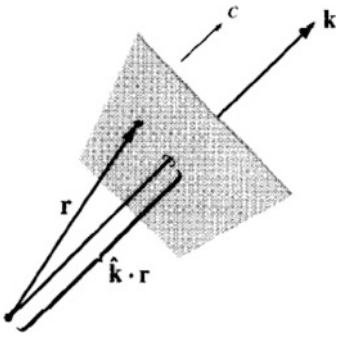


Fig. F.17 Scalar product of $\vec{k} \cdot \vec{r}$

$$\widetilde{\vec{B}}_0 = \frac{k}{\omega}\left(\hat{z} \times \widetilde{\vec{E}}_0\right) \tag{F.21}$$

From this equation it is evident that, $\vec{E}$ and $\vec{B}$ are *in phase and mutually perpendicular*; their (real) amplitudes are related by

$$B_0 = \frac{k}{\omega}E_0 = \frac{1}{c}E_0 \tag{F.22}$$

The fourth of Maxwell's equation, $\nabla \times \vec{B} = \mu_0\varepsilon_0\left(\partial \vec{E}/\partial t\right)$, does not yield an independent condition; it simply reduces to Eq. F.20.

We can easily generalize to monochromatic plane waves traveling in an arbitrary direction. The notation is facilitated by the introduction of **propagation** (or **wave**) **vector** $\vec{k}$ pointing in the direction of propagation, whose magnitude is the wave number k. The scalar product $\vec{k} \cdot \vec{r}$is the appropriate generalization of kzin Fig. F.17, so we have:

$$\begin{aligned} \widetilde{\vec{E}}(\vec{r},t) &= \widetilde{\vec{E}}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}\hat{n} \\ \widetilde{\vec{B}}(\vec{r},t) &= \frac{1}{c}\widetilde{\vec{E}}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}\left(\hat{\vec{k}} \times \hat{n}\right) = \frac{1}{c}\left(\hat{\vec{k}} \times \hat{\vec{E}}\right) \end{aligned} \tag{F.23}$$

where $\hat{n}$ is the polarization vector. Because $\vec{E}$ is transverse,

$$\hat{n} \cdot \hat{\vec{k}} = 0 \tag{F.24}$$

The transversality of $\vec{B}$ follows automatically from Eq. F.23. The actual (real) electric and magnetic fields in a monochromatic plane wave with propagation vector $\vec{k}$ and polarization $\hat{n}$ are [1];

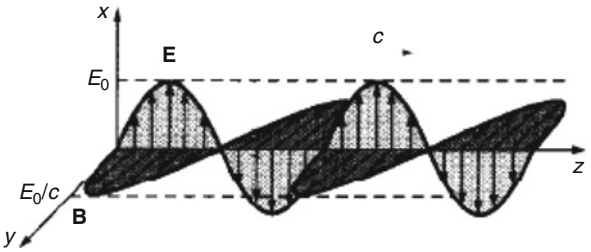


Fig. F.18 Depiction of Example 1 solution

$$\vec{E}(\vec{r},t) = E_0 \cos\left(\vec{k}\cdot\vec{r} - \omega t + \delta\right)\hat{\vec{n}} \tag{F.25}$$

$$\vec{B}(\vec{r},t) = \frac{1}{c}E_0 \cos\left(\vec{k}\cdot\vec{r} - \omega t + \delta\right)\left(\hat{\vec{k}} \times \hat{\vec{n}}\right) \tag{F.26}$$

Example 1: If $\vec{E}$ points in the x-direction, then $\vec{B}$ points in the y-direction Eq. F.21.

Solution: Writing the following

$$\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}\hat{\vec{x}} \quad \text{and} \quad \widetilde{\vec{B}}(z,t) = \frac{1}{c}\widetilde{\vec{E}}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}\hat{\vec{y}}$$

or taking the real part we have (Fig. F.18)

$$\begin{aligned}\vec{E}(z,t) &= E_0 \cos(kz - \omega t + \delta)\hat{\vec{x}} \\ \vec{B}(z,t) &= \frac{1}{c}E_0 \cos(kz - \omega t + \delta)\hat{\vec{y}}\end{aligned} \tag{F.27}$$

Example 2: Write down the (real) electric and magnetic fields for a monochromatic plane

wave of amplitude E_0, frequency ω, and phase angle zero that is (a) traveling in the negative x-direction and polarized in the z-direction; (b) traveling in the direction from the origin to the point (I, I, I), with polarization parallel to the xz-plane. In each case, sketch the wave, and give the explicit Cartesian components of $\vec{k}$ and $\hat{\vec{n}}$.

Solution: we can start with the following for part (a)

$$\vec{k} = \frac{\omega}{c}\hat{\vec{x}} \quad \vec{k}\cdot\vec{r} = \left(-\frac{\omega}{c}\hat{\vec{x}}\right)\cdot\left(x\hat{\vec{x}} + y\hat{\vec{y}} + z\hat{\vec{z}}\right) = \frac{\omega}{c}x \quad \hat{\vec{k}} \times \hat{\vec{n}} = -\hat{\vec{x}} \times \hat{\vec{z}}$$

$$\vec{E}(x,t) = E_0 \cos\left(\frac{\omega}{c}x + \omega t\right)\hat{\vec{z}} \qquad \vec{B}(x,t) = \frac{E_0}{c}\cos\left(\frac{\omega}{c}x + \omega t\right)\hat{\vec{y}}$$

and for part (b) we have the following

$\vec{k} = \frac{\omega}{c}\left(\frac{\hat{x}+\hat{y}+\hat{z}}{\sqrt{3}}\right)$ $\hat{n} = \frac{\hat{x}-\hat{z}}{\sqrt{2}}$ (since $\hat{n}$ is parallel to the xz-plane, it must have the form $\alpha\hat{x} + \beta\hat{z}$; since $\hat{n} \cdot \hat{k} = 0, \beta = -\alpha$; and since it is a unit vector, $\alpha = 1/\sqrt{21}$)

$$\vec{k} \cdot \vec{r} = \frac{\omega}{\sqrt{3}c}\left(\hat{x} + \hat{y} + \vec{z}\right) \cdot \left(x\hat{x} + y\hat{y} + z\hat{z}\right) = \frac{\omega}{\sqrt{3}c}(x + y + x)$$

$\hat{k} \times \hat{n} = \frac{1}{\sqrt{6}}\left|\begin{matrix}\hat{x} & \hat{y} & \hat{z} \\ 1 & 1 & 1 \\ 0 & -1 & \end{matrix}\right| = \frac{1}{\sqrt{6}}\left(-\hat{x} + 2\hat{y} - \hat{z}\right)$, therefore we have the following as final solution for part (b)

$$\vec{E}(x, y, z, t) = E_0 \cos\left[\frac{\omega}{\sqrt{3}c}(x + y + x) - \omega t\right]\left(\frac{\hat{x} - \hat{z}}{\sqrt{2}}\right)$$

$$\vec{B}(x, y, z, t) = E_0 \cos\left[\frac{\omega}{\sqrt{3}c}(x + y + x) - \omega t\right]\left(\frac{-\hat{x} + 2\hat{y} - \hat{z}}{\sqrt{3}}\right)$$

F.6 Boundary Conditions: Reflection & Transmission (Refraction) Dielectric Interface

The important problem of the Reflection and Refraction (Transmission) of a wave at the interface of two media of different dielectric constants is a boundary value problem in which the physics is identical in principle to that involved in time harmonic problems involving conductors. If, for example, one medium is vacuum and a plane wave is incident on a second (dielectric) medium, the incident harmonic of plane wave generates oscillating time harmonic dipoles (or dipolar currents) that produce a field of their own [2]. The strength of the currents is, however, not known in advance and therein lays the essence of problem. Although this problem is a boundary value problem, it is especially simple because of the great symmetry and simple geometry. It turns out that it is solved by a adding to the incident plane waves only two other plane waves, a *reflected* and a *transmitted* (or refracted) one. The geometry is shown Fig. F.20 with the plane $z = 0$ taken as the interface between the two media, labeled 1 and 2.

In terms of the *vector* propagation constant, $\vec{k}_i$, the incident fields are

$$\vec{E}_i = \vec{A}e^{i\vec{k}_i \cdot \vec{r}} \quad \text{and} \quad \vec{B}_i = \left(\vec{k}_i/k\right) \times \vec{E}_i \quad \text{Incident} \tag{F.28}$$

Here $\vec{B}_i$ is derived from the assumed $\vec{E}_i$ by $\nabla \times \vec{E}_i = ik\vec{B}_i$. The normal $\hat{z}$ to the plane $z = 0$ and the vector $\vec{k}_i$ define a *plane of incidence* which can, without loss of generality, be taken to be the x, z-plane as we have done in Fig. F.19. We now postulate the existence of two other plane waves and will show that these suffice to

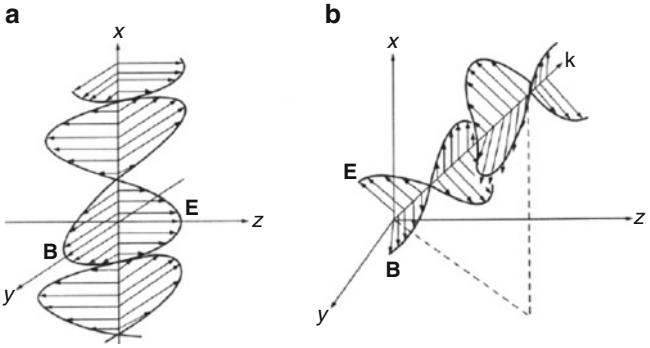


Fig. F.19 Sketch of the Example 2 solution

solve the boundary value problem. These *reflected* and *transmitted* waves, amplitudes $\vec{E}_r$ and $\vec{E}_t$, respectively, are [2]

$$\vec{E}_r = \vec{R}e^{i\vec{k}_r \cdot \vec{r}} \quad \text{and} \quad \vec{B} = \left(\vec{k}_r/k\right) \times \vec{E}_r \quad \text{Reflected} \tag{F.29}$$

$$\vec{E}_t = \vec{T}e^{i\vec{k}_t \cdot \vec{r}} \quad \text{and} \quad \vec{B}_t\Big| = \left(\vec{k}_t/k\right) \times \vec{E}_t \quad \text{Transmitted} \tag{F.30}$$

The vector $\vec{k}_r$ and $\vec{k}_t$ must, for the moment, be considered arbitrary in direction, for although $\vec{k}_i$ is in the x, z-plane, we cannot assume the same a *priori* for $\vec{k}_r$ and $\vec{k}_t$. The magnitudes $k_i \equiv \left|\vec{k}_i\right|$ etc. of the wave vectors are, with n_1 and n_2 the indices of refraction of the two media and $k = \omega/c$,

$$\begin{aligned} k_i &= k_r = n_1 k \\ k_t &= n_2 k \end{aligned} \tag{F.31}$$

Considering the boundary conditions which states the tangential components of $\vec{E}$ and $\vec{B}$ are continuous we have to satisfy the following two steps;

First: If the tangential components of the three fields in Eqs. F.29, F.30 and F.31 are to be matched at $z = 0$, it also should be clear that the spatial dependence given by the exponents must be identical. However, this is a necessary but not a sufficient condition.

Second: The vector coefficients $\vec{A}$, $\vec{R}$, and $\vec{T}$ must be determined.

The first condition, that the spatial variation of the three fields must be identical at $z = 0$, leads to the following equation.

$$\left(\vec{k}_i \cdot \vec{r}\right)_{z=0} = \left(\vec{k}_r \cdot \vec{r}\right)_{z=0} = \left(\vec{k}_t \cdot \vec{r}\right)_{z=0} \tag{F.32}$$

The first equality in Eq. F.32 yields $k_{ix}x = k_{rx}x + k_{ry}y$. For this condition to hold for all x and y, we must have $k_{ry} = 0$, showing that $\vec{k}_r$ lies in the plane of incidence and also that $k_{ix}x = k_{rx}x$. Similarly, from the second equality in Eq. F.32, $\vec{k}_t$ must also lie in this plane, so that $\vec{k}_i$, $\vec{k}_r$, $\vec{k}_t$ are all coplanar. Moreover, from the geometry of Fig. F.19, we have from $k_{ix} = k_{rx}$ that $k_i \sin\theta_i = k_r \sin\theta_r$. Or since $k_i = k_r$, then we can write the following;

$$\sin\theta_i = \sin\theta_r \tag{F.33}$$

The *angle of incidence equals the angle of reflection*. Similarly, the equality of k_{ix} and k_{ix} yields

$$k_i \sin\theta_i = k_t \sin\theta_t$$

Or using Eq. F.31, we have

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_2}{n_1} \tag{F.34}$$

Equation F.34 is known as *Snell's law* of refraction. The condition of Eqs. F.33 and F.34 are quite general ones that are independent of the detailed vectorial nature of the wave field.. They hold for reflection and refraction of scalar waves. These conditions by themselves do not guarantee the continuity of tangential $\vec{E}$ and $\vec{B}$ across the boundary. To satisfy these conditions, more specification of polarization of the fields should be analyzed [2]. For convenient we consider the general the general case of arbitrary incident polarization as a linear combination of a wave with polarization perpendicular to the plane of incidence and one with polarization parallel to this plane. The reflected and transmitted waves will then be similarly polarized. These two cases are sketched as parts (a) and (b) in Fig. F.20.

Consider first the case of $\vec{E}$ perpendicular to the plane of incidence, i.e., in the y-direction. The vector coefficients $\vec{A}$, $\vec{R}$, and $\vec{T}$ becomes scalar ones, with subscript $\perp$ to denote this case.

$$\begin{aligned} E_{i\perp} &= A_\perp e^{i\vec{k}_i \cdot \vec{r}} \\ E_{r\perp} &= A_\perp e^{i\vec{k}_r \cdot \vec{r}} \\ E_{t\perp} &= A_\perp e^{i\vec{k}_t \cdot \vec{r}} \end{aligned} \tag{F.35}$$

Then from the continuity of tangential $\vec{E}$ at the boundary, we have

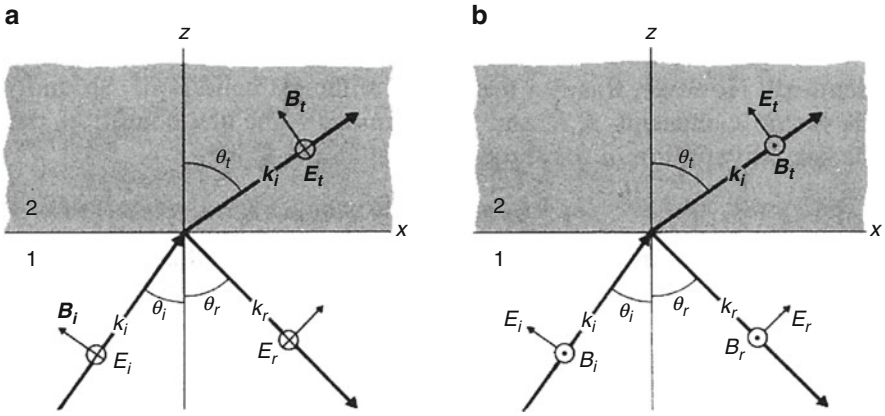


Fig. F.20 Reflection and Refraction at a dielectric interface. (**a**): Electric vector perpendicular to plane of incidence. (**b**): Electric vector parallel to plane of incidence

$$A_\perp + R_\perp = T_\perp \tag{F.36}$$

The condition on tangential $\vec{B}$ becomes, with $B_{ix} = -A_\perp \cos\theta_i, B_{rx} = R_\perp \cos\theta_r$, and $B_{tx} = -T_\perp \cos\theta_t$, also considering that $\cos\theta_i = -\theta_r$ we then have:

$$n_1(A_\perp - R_\perp)\cos\theta_i = n_2 T_\perp \cos\theta_t \tag{F.37}$$

Solving Eqs. F.36 and F.37 for the ratios $R_\perp/A_\perp$ and $T_\perp/A_\perp$, we find, using Snell's law, that

$$\begin{aligned} \frac{R_\perp}{A_\perp} &= \frac{1 - \dfrac{\tan\theta_i}{\tan\theta_r}}{1 + \dfrac{\tan\theta_i}{\tan\theta_r}} = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \\ \frac{T_\perp}{A_\perp} &= \frac{2}{1 + \dfrac{\tan\theta_i}{\tan\theta_r}} = \frac{2\cos\theta_i \sin\theta_t}{\sin(\theta_i + \theta_t)} \end{aligned} \tag{F.38}$$

For the second case, incident wave polarized parallel to the plane of incidence, we use $A_{||} = |\vec{A}|$ in Eq. F.28 and, similarly $R_{||} = |\vec{R}|$ and $T_{||} = |\vec{T}|$ in Eqs. F.29 and F.30. The boundary condition yield

$$\begin{aligned} \cos\theta_i (A_{||} - R_{||}) &= \cos\theta_i T_{||} \\ n_1 (A_{||} - R_{||}) &= n_2 T_{||} \end{aligned}$$

These equations lead to the results

$$\frac{R_{||}}{A_{||}} = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)}$$
$$\frac{T_{||}}{A_{||}} = \frac{2\cos\theta_i \sin\theta_t}{\sin(\theta_i + \theta_t)\cos(\theta_i - \theta_t)} \tag{F.39}$$

There are two phenomena worthy of note in connection with the above discussion. Consider the case of polarization in the plane of incidence. We see from Eq. F.39 that $R_{||}/A_{||}$ will be zero for $\theta_i + \theta_t = \pi/2$. Putting this condition into Snell's law, using $\sin\theta_t = \sin(\pi/2 - \theta_i) = \cos\theta_i$, we see that the angle of incidence θ_B (called the *Brewster angle*) for which this happens is defined by

$$\tan\theta_B = \frac{n_2}{n_1} \tag{F.40}$$

If wave with arbitrary polarization is incident on a dielectric interface, it can be considered to be a linear combination of a wave polarized parallel to, and a wave polarized perpendicular to, the plane of incident. At the Brewster angle, the parallel component will not be reflected so that the reflected wave will be plane polarized in a plane perpendicular to the plane of incidence. This effect can then be made the basis of a device for polarizing an unpolarized beam of radiation.

The second phenomenon is that of *total internal reflection*. In either (a) or (b) of Fig. F.4, suppose that the index n_1 is greater than n_2. Then, from Snell's law, $\sin\theta_t = (n_1/n_2)\sin\theta_i$ is always greater than θ_i.There will then be some value of θ_i, call it θ_{int} for which $\theta_t = \pi/2$; this angle is defined by

$$\sin\theta_{\text{int}} = \frac{n_2}{n_1}$$

Since in general

$$\vec{E}_t = \vec{T}e^{ikn_2(x\sin\theta_t + z\cos\theta_t)} \tag{F.41}$$

for $\theta_t = \pi/2$ there is no wave in the second medium; the z-dependence vanishes. Now if $\theta_i > \theta_{\text{int}}$, then $\sin\theta_t$, is larger than unity from Snell's law and, as a consequence, $\cos\theta_t$ is imaginary

$$\cos\theta_t = \sqrt{1 - \left(\frac{n_1}{n_2}\right)\sin^2\theta_i} = i\sqrt{\left(\frac{n_1}{n_2}\right)^2 \sin^2\theta_i - 1}$$

Equation F.32 then becomes

$$\vec{E}_t = \vec{T} e^{-kn_2 z\sqrt{(n_1/n_2)\sin^2\theta_i - 1}} e^{ikn_2 x \sin\theta_t}$$

This corresponds to a wave which is exponentially attenuated as a function of z, and which propagates as a function of x with propagation constant $kn_2 \sin\theta_t$. Such a wave is the prototype of a surface wave and reader can see further discussion and more details in Reference 2 of this Appendix *2*.

The interesting question is this: What happens when a wave passes from one transparent medium into another—air to water, say, or glass to plastic? As in the case of waves on a string, we expect to get a reflected wave and a transmitted wave. The details depend on the exact nature of the electrodynamics boundary conditions F.1.

$$\left.\begin{array}{ll} \text{(i)}\,\varepsilon_1 E_1^{\perp} = \varepsilon_2 E_x^{\perp} & \text{(iii)}\,E_1^{||} = E_2^{||} \\ \text{(ii)}\,B_1^{\perp} = B_2^{\perp} & \text{(iv)}\,\dfrac{1}{\mu_1} B_1^{||} = \dfrac{1}{\mu_2} B_2^{||} \end{array}\right\} \tag{F.42}$$

These equations relate the electric and magnetic fields just to the left and just to the right of the interface between two linear media. In the following sections we use them to deduce the laws governing reflection and refraction of electromagnetic waves.

F.7 Electromagnetic Waves in Matter

Some solutions to Maxwell's Equations have been already discussed in previous sections. The present section extends the treatment of electromagnetic waves. Since most regions of interest are free of charge, it will be assumed that charge density $\rho = 0$. Moreover, linear isotropic (invariant with respect to direction) materials will be assumed, with the following relationships:

$$\begin{cases} \vec{D} = \varepsilon \vec{E} \\ \vec{B} = \mu \vec{H} \\ \vec{J} = \sigma \vec{E} \end{cases} \tag{F.43}$$

Where

$\vec{D}$ = Flux Density (c/m^2)
$\vec{E}$ = Electric Field (N/C)
ε = Permittivity of the Medium (C^2/N. m^2) or Equivalent (F/m)
$\vec{B}$ = Magnetic Field (T)
$\vec{H}$ = Magnetic Filed Strength (A/m)

μ = Mobility within Materials ($m^2/V.s$)
$\vec{J}$ = Current Density (A/m^2)
σ = Conductivity of Materials (S/m)
ρ = Charge Density (C/m^3)
where
A = Ampere
C = Coulomb
N = Newton
F = Farad
S = Siemens
T = Tesla

F.7.1 Electromagnetic Waves in Matter

Inside matter, but in regions where there is no *free* charge or *free current*, Maxwell's equations become

$$\left.\begin{array}{ll} \text{(i)} \nabla \cdot \vec{D} = 0 & \text{(iii)} \nabla \times \vec{E} = -\dfrac{\partial \vec{B}}{\partial t} \\ \text{(ii)} \nabla \cdot \vec{B} = 0 & \text{(iv)} \nabla \times \vec{H} = -\dfrac{\partial \vec{D}}{\partial t} \end{array}\right\} \tag{F.44}$$

If the medium is *linear*

$$\vec{D} = \varepsilon \vec{E} \qquad \text{and} \qquad \vec{H} = \frac{1}{\mu}\vec{B} \tag{F.45}$$

and homogeneous (so ε and μ do not vary from point to point), Maxwell's equations reduce to

$$\left.\begin{array}{ll} \text{(i)} \nabla \cdot \vec{E} = 0 & \text{(iii)} \nabla \times \vec{E} = -\dfrac{\partial \vec{B}}{\partial t} \\ \text{(ii)} \nabla \cdot \vec{B} = 0 & \text{(iv)} \nabla \times \vec{B} = \mu\varepsilon \dfrac{\partial \vec{E}}{\partial t} \end{array}\right\} \tag{F.46}$$

which remarkably differ from the vacuum analogs Eq. F.1 only in the replacement of $\mu_0\varepsilon_0$ by $\mu\varepsilon$. This mathematically is obvious, yet the physical implications are astonishing [4]. As the wave passes through, the fields busily polarized and magnetize all the molecules, and the resulting (oscillating) dipoles create their own electric and magnetic fields. These combine with the original fields in such a way as to create a *single* wave with the same frequency but a different speed. This

extraordinary conspiracy is responsible for the phenomenon of **transparency**. It is a distinctly *nontrivial* consequence of the *linearity* of the medium [4]. Looking at set of Eq. of 45 it is evident electromagnetic waves propagate through a linear homogeneous medium at a speed

$$v = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c}{n} \tag{F.47}$$

where

$$n \equiv \sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}} \tag{F.48}$$

is the index of refraction of the material. For most materials, μ is very close to μ_0, so

$$n \cong \sqrt{\varepsilon_r} \tag{F.49}$$

where ε_r is the **dielectric constant** or also known as **relative permittivity** [1] and is equal to $\varepsilon_r = \varepsilon/\varepsilon_0$. Since ε_r is almost always greater than 1, light travels *more slowly* through matter—a fact that is well known from optics [1].

F.7.2 Reflection and Transmission at Normal Incidence

Suppose the xy-plane forms the boundary between two linear media. A plane wave of frequency ω, traveling in the z-direction and polarized in the x-direction, approaches the interface from the left (Fig. F.21):

$$\left.\begin{aligned} \widetilde{\vec{E}}_1(z,t) &= \widetilde{E}_{0_I} e^{i(k_1 z - \omega t)}\,\hat{\vec{x}} \\ \vec{B}_I(z,t) &= \frac{1}{v_1}\widetilde{E}_{0_I} e^{i(k_1 z - \omega t)}\,\hat{\vec{y}} \end{aligned}\right\} \tag{F.50}$$

It is given rise to a reflected wave

$$\left.\begin{aligned} \widetilde{\vec{E}}_R(z,t) &= \widetilde{E}_{0_R} e^{i(k_1 z - \omega t)}\,\hat{\vec{x}} \\ \vec{B}_R(z,t) &= \frac{1}{v_1}\widetilde{E}_{0_R} e^{i(-k_1 z - \omega t)}\,\hat{\vec{y}} \end{aligned}\right\} \tag{F.51}$$

which travels back to the left in medium (F.1), and a transmitted waves

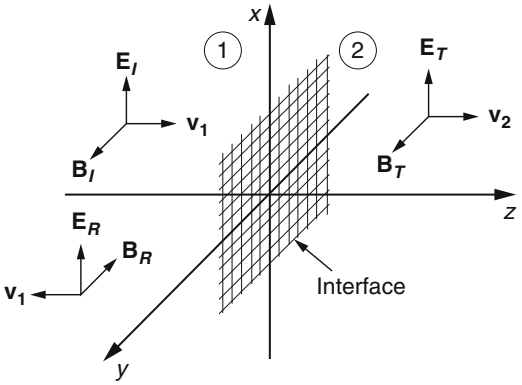


Fig. F.21 Normal wave incidence [1]

$$\left.\begin{aligned}\widetilde{\vec{E}}_T(z,t) &= \widetilde{E}_{0_T} e^{i(k_2 z-\omega t)}\,\hat{x}\\ \vec{B}_T(z,t) &= \frac{1}{v_2}\widetilde{E}_{0_T} e^{i(k_2 z-\omega t)}\,\hat{y}\end{aligned}\right\} \tag{F.52}$$

which continues on the right in medium (F.2). Note the minus sign in $\vec{B}_R$, as required by Eq. F.23 or, if you prefer, by the fact that the Poynting vector aims in the direction of propagation.

At $z = 0$, the combined fields on the left, $\widetilde{\vec{E}}_I + \widetilde{\vec{E}}_R$ and $\widetilde{\vec{B}}_I + \widetilde{\vec{B}}_R$, must join the fields on the right, $\widetilde{\vec{E}}_T$ and $\widetilde{\vec{B}}_T$, in accordance with the boundary conditions Eq. F.42. In this case there are no components perpendicular to the surface, so (i) and (ii) are trivial. However, (iii) requires that

$$\widetilde{E}_{0_I} + \widetilde{E}_{0_R} = \widetilde{E}_{0_T} \tag{F.53}$$

while (iv) says

$$\frac{1}{\mu_1}\left(\frac{1}{v_1}\widetilde{E}_{0_I} - \frac{1}{v_1}\widetilde{E}_{0_R}\right) = \frac{1}{\mu_2}\left(\frac{1}{v_2}\widetilde{E}_{0_T}\right) \tag{F.54}$$

or

$$\widetilde{E}_{0_I} - \widetilde{E}_{0_R} = \beta\widetilde{E}_{0_T} \tag{F.55}$$

where

$$\beta \equiv \frac{\mu_1 v_1}{\mu_2 v_2} = \frac{\mu_1 n_2}{\mu_2 n_1} \tag{F.56}$$

Equations F.53 and F.55 are easily solved for the outgoing amplitudes, in terms of the incident amplitude:

$$\vec{E}_{0_R} = \left(\frac{1-\beta}{1+\beta}\right)\widetilde{E}_{0_I} \quad \text{and} \quad \widetilde{E}_{0_T} = \left(\frac{2}{1+\beta}\right)\widetilde{E}_{0_I} \tag{F.57}$$

These results are strikingly similar to the ones for waves on a string. Indeed, if the perrnittivities μ *are* close to their values in vacuum (as, remember, they are for most media), then $\beta = \dfrac{v_1}{v_2}$, and we have [1]

$$\widetilde{E}_{0_R} = \left(\frac{v_2 - v_1}{v_2 + v_1}\right)\widetilde{E}_{0_I} \quad \text{and} \quad \widetilde{E}_{0_r} = \left(\frac{2v_2}{v_2 + v_1}\right)\widetilde{E}_{0_I} \tag{F.58}$$

In that case, as before, the reflected wave is in *phase* (right side up) if $v_2 > v_1$ and *out of phase* (upside down) if $v_2 < v_1$; the real amplitudes are related by

$$E_{0_R} = \left(\frac{v_2 - v_1}{v_2 + v_1}\right)E_{0_I} \quad \text{and} \quad E_{0_r} = \left(\frac{2v_2}{v_2 + v_1}\right)E_{0_I} \tag{F.59}$$

or, in terms of the indices of refraction [1],

$$E_{0_R} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)E_{0_I} \quad \text{and} \quad E_{0_T} = \left(\frac{2n_1}{n_1 + n_2}\right)E_{0_I} \tag{F.60}$$

According to definition of intensity (average power per unit area) that is given by D. Griffiths [1] as follows, we should ask what fraction of the incident energy is reflected, and what fraction is transmitted?

$$I = \frac{1}{2}\varepsilon v E_0^2 \tag{F.61}$$

If again $\mu_1 = \mu_2 = \mu_0$, then the ratio of the reflected intensity to the incident intensity is

$$R \equiv \frac{I_R}{I_I} = \left(\frac{E_{0_R}}{E_{0_I}}\right)^2 = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \tag{F.62}$$

whereas the ratio of the transmitted intensity to the incident intensity is

$$T \equiv \frac{I_T}{I_I} = \frac{\varepsilon_2 v_2}{\varepsilon_1 v_1}\left(\frac{E_{0_T}}{E_{0_I}}\right)^2 = \frac{4n_1 n_2}{(n_1 + n_2)^2} \tag{F.63}$$

R is called the **Reflection coefficient** and T the **Transmission coefficient**; they measure the fraction of the incident energy that is reflected and transmitted, respectively. Notice that [1]

$$R + T = 1 \tag{F.64}$$

as conservation of energy, of course, requires. For instance, when light passes from air ($n_1 = 1$)into glass ($n_2 = 1.5$), $R = 0.04$ and $T = 0.96$. Not surprisingly, most of the light is transmitted [1].

Example 1: Calculate the exact reflection and transmission coefficients, without assuming $\mu_1\mu_2 = \mu_0$. Confirm $R + T = 1$.

Solution: From Eq. F.62 we have $R = \left(\frac{E_{0_R}}{E_{0_I}}\right)^2$ substituting from Eq. F.57 results where $\beta \equiv \frac{\mu_1 v_1}{\mu_2 v_2}$ and using Eq. F.63 $T = \frac{\varepsilon_2 v_2}{\varepsilon_1 v_1}\left(\frac{E_{0_T}}{E_{0_I}}\right)^2$ results that $T = \beta\left(\frac{2}{1+\beta}\right)^2$ which is Eq. F.57. Note that $\frac{\varepsilon_2 v_2}{\varepsilon_1 v_1} = \frac{\mu_1}{\mu_2}\frac{\varepsilon_2 v_2}{\varepsilon_1 v_1}\frac{v_2}{v_1} = \frac{\mu_1}{\mu_2}\left(\frac{v_1}{v_2}\right)^2 = \frac{\mu_1 v_1}{\mu_2 v_2} = \beta$

$$\begin{aligned} R &= \frac{2}{(1+\beta)^2}\left[4\beta + \left(1-\beta^2\right)\right] \\ &= \frac{1}{(1+\beta)^2}\left(4\beta + 1 - 2\beta + \beta^2\right) \\ &= \frac{1}{(1+\beta)^2}\left(1 + 2\beta + \beta^2\right) = 1 \end{aligned}$$

Example 2: In writing Eqs. F.51 and F.52, assumption was the reflected and transmitted waves have the same *polarization* as the incident wave—along the x–direction. Prove that this must be so. [Hint: Let the polarization vector of the transmitted and reflected wave be $\hat{n}_T = \cos\theta_T \hat{x} + \sin\theta_T \hat{y}$ and $\hat{n}_R = \cos\theta_R \hat{x} + \sin\theta_R \hat{y}$ and prove from the boundary condition that $\theta_T = \theta_R = 0$]

Solution: Equation F.43 is replaced by $\widetilde{E}_{0_I}\hat{x} + \widetilde{E}_{0_R}\hat{n} = \widetilde{E}_{0_T}\hat{n}_T$ and Eq. F.55 $\widetilde{E}_{0_I}\hat{y} - \widetilde{E}_{0_R}\left(\hat{z} \times \hat{n}_R\right) = \beta\widetilde{E}_{0_T}\left(\hat{z} \times \hat{n}_T\right)$. The y component of the first equation is $\widetilde{E}_{0_R}\sin\theta_R = \widetilde{E}_{0_T}\sin\theta_T$; the x component of the second is $\widetilde{E}_{0_R}\sin\theta_R = -\beta\widetilde{E}_{0_T}\sin\theta_T$. Comparing these two, we conclude that $\sin\theta_R = \sin\theta_T = 0$, and hence $\theta_T = \theta_R = 0$.

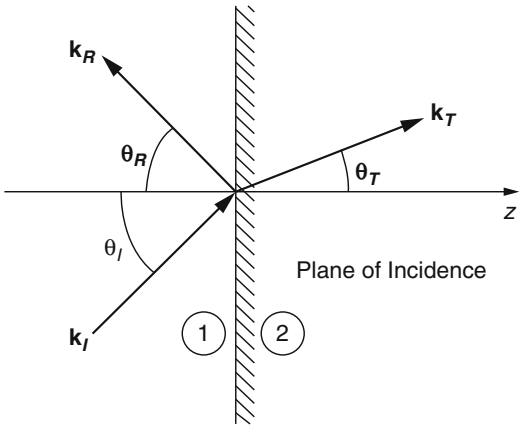


Fig. F.22 Incidence at oblique angle [1]

F.7.3 Reflection and Transmission at Normal Incidence

In the last section Reflection and Transmission was treated at *Normal* incidence—that is, when the incoming wave hits the interface head-on at 90 degree incident angle [1]. The more general case of *Oblique* incidence, in which the incoming wave meets the boundary at an arbitrary angle θ_I also is treated by David Griffiths reference 1 of this appendix. We just write the result of his conclusion here and we encourage the reader to refer to his book section 9.3.3 page 386 for more details [1]. Although the angle of incidence $\theta_I = 0$ is special case of the oblique incidence, but we did treat that separately for some cases that we are dealing with high power laser interaction with matter in this book. The treatment by Griffiths is done based on monochromatic plane wave with such arbitrary angle of incidence. See Fig. F.22 here.

Suppose that a monochromatic plane wave (incident wave) approaches from the left as follow

$$\widetilde{\vec{E}}_I(\vec{r},t) = \widetilde{\vec{E}}_{0_I} e^{i\left(\vec{k}_I \cdot \vec{r} - \omega t\right)} \quad \text{and} \quad \widetilde{\vec{B}}_I(\vec{r},t) = \frac{1}{v_1}\left(\hat{\vec{k}}_I \times \widetilde{\vec{E}}_I\right) \tag{F.65}$$

And giving rise to a reflected wave of the form in Eq. F.66

$$\widetilde{\vec{E}}_R(\vec{r},t) = \widetilde{\vec{E}}_{0_R} e^{i\left(\vec{k}_R \cdot \vec{r} - \omega t\right)} \quad \text{and} \quad \widetilde{\vec{B}}_R(\vec{r},t) = \frac{1}{v_1}\left(\hat{\vec{k}}_R \times \widetilde{\vec{E}}_R\right) \tag{F.66}$$

and a transmitted set of wave Eq. of 66;

$$\widetilde{\vec{E}}_T(\vec{r},t) = \widetilde{\vec{E}}_{0_T} e^{i\left(\vec{k}_T \cdot \vec{r} - \omega t\right)} \quad \text{and} \quad \widetilde{\vec{B}}_T(\vec{r},t) = \frac{1}{v_2}\left(\hat{\vec{k}}_T \times \widetilde{\vec{E}}_T\right) \tag{F.67}$$

All three waves have the same frequency ω—that is determined once for all at the source (the laser beam, flashlight or whatever, that produces the incident beam). The three wave numbers are related per Eq. F.11;

$$k_I v_1 = k_R v_1 = k_T v_2 = \omega \qquad \text{or} \qquad k_I = k_R = \frac{v_2}{v_1} k_T = \frac{n_1}{n_2} k_T \tag{F.68}$$

The combined fields in medium (F.1), $\widetilde{\vec{E}}_I + \widetilde{\vec{E}}_R$ and $\widetilde{\vec{B}}_I + \widetilde{\vec{B}}_R$, must now be joined to the fields $\widetilde{\vec{E}}_T$ and $\widetilde{\vec{B}}_T$ in medium (F.2), using the boundary conditions Eq. F.42. These all share the generic structure. *Boundary conditions should hold at all points on the plane, and for all times for all above three plane wave equations when* $z = 0$. Otherwise, a slight change in x, would destroy the equality (See Example 1 below) [1]. The time factors in this case *already* equal, in fact we can regard this as an independent confirmation that the transmitted and reflected frequencies must match the incident frequency. So for spatial case we have the following relation;

$$\vec{k}_I \cdot \vec{r} = \vec{k}_R \cdot \vec{r} = \vec{k}_T \cdot \vec{r} \quad \text{when} \quad z = 0 \tag{F.69}$$

or more explicitly,

$$x(k_I)_x + y(k_R)_x + y(k_R)_y = x(k_T)_x + y(k_T)_y \tag{F.70}$$

for all z and y. But Eq. F.70 can *only* hold if the components are separately equal, for if $x = 0$, we get

$$(k_I)_y = (k_R)_y = (k_T)_y \tag{F.71}$$

While $y = 0$ gives

$$(k_I)_x = (k_R)_x = (k_T)_x \tag{F.72}$$

We may as well orient our axes so that $\vec{k}$ lies in the xy plane (i.e. $(k_I)_y = 0$); according to Eq. F.71, so too will $\vec{k}_R$ and $\vec{k}_T$. *Conclusion*:

First Law: The incident, reflected, and transmitted wave vectors form a plane (called the **plane of incidence**), which also includes the normal to the surface (here, the z-axis).

Meanwhile, Eq. F.72 implies that

$$k_I \sin\theta_I = k_R \sin\theta_R = k_T \sin\theta_T \tag{F.73}$$

where θ_I is the **angle of incidence**, θ_R is the **angle of reflection**, and θ_T is the **angle of transmission**, more commonly known as the **angle of refraction**, all of them measured with respect to the normal (Fig. F.22). In view of Eq. F.68, then,

Second Law: The angle of incidence is equal to the angle of reflection,

$$\theta_I = \theta_T \tag{F.74}$$

This is the **law of reflection**.
As for the transmitted angle,
Third Law:

$$\frac{\sin\theta_T}{\sin\theta_I} = \frac{n_1}{n_2} \tag{F.75}$$

This is the **law of refraction**, or **Snell's law**.

These are the three fundamental laws of geometrical optics. It is remarkable how little actual electrodynamics went into them: we have yet to invoke any specific boundary conditions. Now that we have taken care of the exponential factors—they cancel, given Eq. F.69 the boundary conditions Eq. F.42 become [1]:

$$\left.\begin{aligned}
&\text{(i)} \quad \varepsilon_1\left(\tilde{\vec{E}}_{0_I} + \tilde{\vec{E}}_{0_R}\right)_z = \varepsilon_2\left(\tilde{\vec{E}}_{0_T}\right)_z \\
&\text{(ii)} \quad \left(\tilde{\vec{B}}_{0_I} + \tilde{\vec{B}}_{0_R}\right)_z = \left(\tilde{\vec{B}}_{0_T}\right)_z \\
&\text{(iii)} \quad \left(\tilde{\vec{E}}_{0_I} + \tilde{\vec{E}}_{0_R}\right)_{x,y} = \left(\tilde{\vec{E}}_{0_T}\right)_{x,y} \\
&\text{(iv)} \quad \frac{1}{\mu_1}\left(\tilde{\vec{B}}_{0_I} + \tilde{\vec{B}}_{0_R}\right)_{x,y} = \frac{1}{\mu_2}\left(\tilde{\vec{B}}_{0_T}\right)_{x,y}
\end{aligned}\right\} \tag{F.76}$$

where $\tilde{\vec{B}}_0 = (1/v)\hat{\vec{k}} \times \tilde{\vec{E}}_0$ in each case. The last two (iii) and (iv) represent pairs of equations, one for the x–direction and one for the y–direction.

Suppose that the polarization of the incident wave is *parallel* to the plane of incidence (the xz plane in Fig. F.23 below); it follows (See Example 2, Section 1.6.2 of this appendix in above) that the reflected and transmitted waves are also polarized in this plane. We will present that in an example to analyze the case of polarization perpendicular to the plane of incidence (See Example 2 below.) Then (i) reads;

$$\varepsilon_1\left(-\tilde{\vec{E}}_{0_I}\sin\theta_I + \tilde{\vec{E}}_{0_R}\sin\theta_R\right) = \varepsilon_2\left(-\tilde{\vec{E}}_{0_T}\sin\theta_T\right) \tag{F.77}$$

(ii) adds nothing ($0 = 0$), since the magnetic fields have no z components; (iii) becomes [1];

$$\tilde{\vec{E}}\cos\theta_I + \tilde{\vec{E}}_{0_R}\cos\theta_R = \tilde{\vec{E}}_{0_T}\cos\theta_T \tag{F.78}$$

and (iv) says

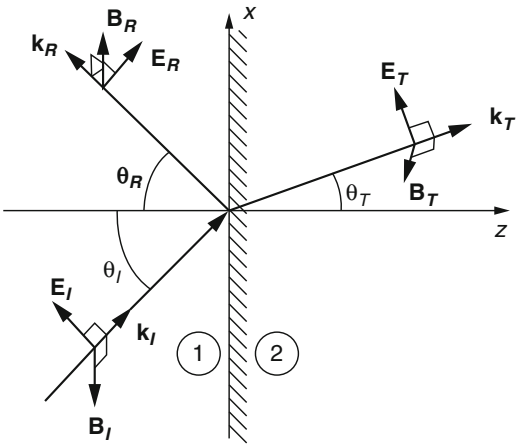


Fig. F.23 Incident Oblique Wave

$$\frac{1}{\mu_1 v_1}\left(\tilde{\vec{E}}_{0_I} - \tilde{\vec{E}}_{0_R}\right) = \frac{1}{\mu_2 v_2}\left(\tilde{\vec{E}}_{0_T}\right) \tag{F.79}$$

Given the laws of reflection and refraction, Eqs. F.76 and F.79 both reduce to

$$\tilde{\vec{E}}_{0_I} - \tilde{\vec{E}}_{0_R} = \beta \tilde{\vec{E}}_{0_T} \tag{F.80}$$

Where again

$$\beta \equiv \frac{\mu_1 v_1}{\mu_2 v_2} = \frac{\mu_1 n_2}{\mu_2 n_1} \tag{F.81}$$

Equation F.78 says

$$\tilde{\vec{E}}_{0_I} + \tilde{\vec{E}}_{0_R} = \alpha \tilde{\vec{E}}_{0_T} \tag{F.82}$$

where again

$$\alpha = \frac{\cos\theta_T}{\cos\theta_I} \tag{F.83}$$

Solving Eqs. F.80 and F.82 for the reflected and transmitted amplitudes, we obtain [1];

$$\begin{aligned} \tilde{\vec{E}}_{0_R} &= \left(\frac{\alpha - \beta}{\alpha + \beta}\right)\tilde{\vec{E}}_{0_I} \\ \tilde{\vec{E}}_{0_T} &= \left(\frac{2}{\alpha + \beta}\right)\tilde{\vec{E}}_{0_I} \end{aligned} \tag{F.84}$$

These are known as Fresnel's equations, for the case of polarization in the plane of incidence. (There are two other Fresnel equations, giving the reflected and transmitted amplitudes when the polarization is perpendicular to the plane of incidence-see Prob. 9.16.). Notice that the transmitted wave is always in phase with the incident one; the reflected wave is either in phase ("right side up"), if $\alpha > \beta$, or 180° out of phase ("upside down"). if $\alpha < \beta$.

Note that the, there is an unavoidable ambiguity in the phase of the reflected wave. The convention that David Griffiths [1] has used in his book under Chapter 9, which is reflecting here and adopted in Fig. F.23 with $\widetilde{\vec{E}}_R$ positive "upward" is consistent with some, but not all, of the standard optics text. Changing the sign of the polarization vector is equivalent to a 180° phase shift.

The amplitudes of the transmitted and reflected waves depend on the angle of incidence α because a is a function of θ_I:

$$\alpha = \sqrt{\frac{1 - \sin^2\theta_T}{\cos\theta_I}} = \frac{\sqrt{1 - [(n_1/n_2)\sin\theta_I]^2}}{\cos\theta_I} \tag{F.85}$$

In the case of normal incidence ($\theta_I = 0$), $\alpha = 1$, and we recover Eq. F.57. At grazing incidence ($\theta_I = 90$), α diverges, and the wave is totally reflected (a fact that is painfully familiar to anyone who has driven at night on a wet road). Interestingly, there is an intermediate angle, θ_B (called **Brewster's angle**), at which the reflected wave is completely extinguished [1]. Because waves polarized perpendicular to the plane of incidence exhibit no corresponding quenching of the reflected component, an arbitrary beam incident at Brewster's angle yields a reflected beam that is totally polarized parallel to the interface. That's why Polaroid glasses, with the transmission axis vertical, help to reduce glare off a horizontal surface [1].

According to Eq. F.84, this occurs when $\alpha = \beta$, or

$$\sin^2\theta_B = \frac{1 - \beta^2}{(n_1/n_2)^2 - \beta^2} \tag{F.86}$$

For the typical case $\mu_1 \cong \mu_2$, so $\beta \cong n_2/n_1$, $\sin^2\theta_B \cong \beta^2/(1+\beta^2)$, and hence

$$\tan\theta_B \cong \frac{n_2}{n_1} \tag{F.87}$$

Figure F.24 shows a plot of the transmitted and reflected amplitudes as functions of eI, for light incident on glass ($n_2 = 1.5$) from air ($n_1 = 1.0$). (On the graph, a negative number indicates that the wave is 180° out of phase with the incident beam—the amplitude itself is the absolute value.)

The power per unit area striking the interface is $\vec{S} \cdot \hat{z}$. Thus the incident intensity is

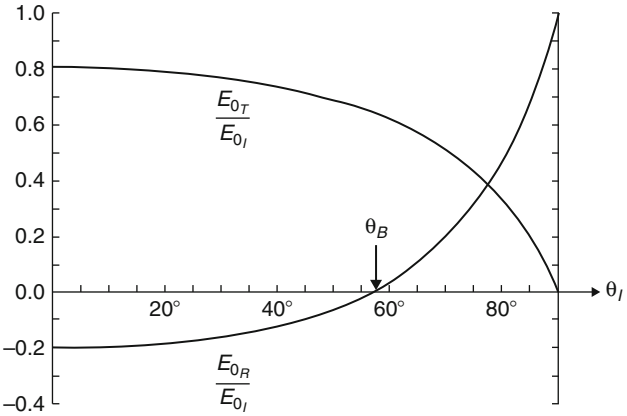


Fig. F.24 Plot of the transmitted and reflected amplitude [1]

$$I_I = \frac{1}{2}\varepsilon_1 v_1 E_{0_I}^2 \cos\theta_I \tag{F.88}$$

while the Reflected and Transmitted intensities are

$$I_I = \frac{1}{2}\varepsilon_1 v_1 E_{0_I}^2 \cos\theta_I \quad \text{and} \quad I_T \frac{1}{2}\varepsilon_2 v_2 E_{0_I}^2 \cos\theta_T \tag{F.89}$$

(The cosines are there because we are talking about the average power per unit area of interface, and the *interface* is at an angle to the wave front.) [1]. The reflection and transmission coefficients for waves polarized parallel to the plane of incidence are

$$R \equiv \frac{I_R}{I_I} = \left(\frac{E_{0_R}}{E_{0_I}}\right)^2 = \left(\frac{\alpha - \beta}{\alpha + \beta}\right)^2 \tag{F.90}$$

$$T \equiv \frac{I_T}{I_I} = \frac{\varepsilon_2 v_2}{\varepsilon_1 v_1}\left(\frac{E_{0_T}}{E_{0_I}}\right)^2 \frac{\cos\theta_T}{\cos\theta_I} = \alpha\beta\left(\frac{2}{\alpha + \beta}\right)^2 \tag{F.91}$$

They are plotted as functions of the angle of incidence in Fig. F.25 (for the air/glass interface). R is the fraction of the incident energy that is reflected-naturally, it goes to zero at Brewster's angle; T is the fraction transmitted—it goes to 1 at θ_B. Note that $R + T = 1$ as required by conservation of energy: the energy per unit time *reaching* a particular patch of area on the surface is equal to the energy per unit time *leaving* the patch [1].

Example 1: Suppose $Ae^{iax} + Be^{ibx} - Ce^{icx}$, for some nonzero constant A, B, C, a, b, c, and for all x. Prove that $a = b = c$ and $A + B = C$.

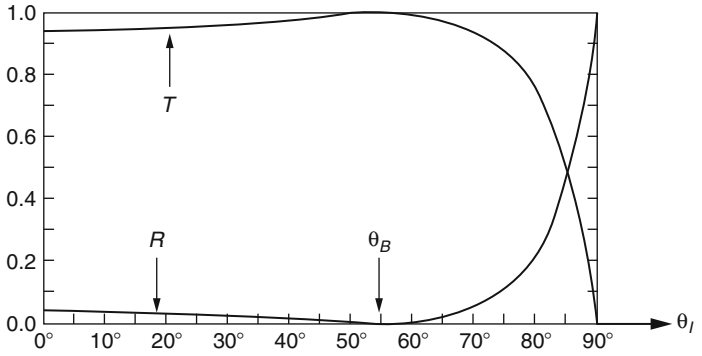


Fig. F.25 Plot of reflection and transmission coefficient for polarized waves

Solution: $Ae^{iax} + Be^{ibx} = Ce^{icx}$, for all x, so (using $x = 0$), $A + B = C$.

Differentiate: $iaAe^{iax} + ibBe^{ibx} = icCe^{icx}$, so (using $x = 0$), $aA + bB = cC$.

Differentiating again: $-a^2Ae^{iax} - b^2Be^{ibx} = -c^2Ce^{icx}$, so (using $x = 0$), $a^2A + b^2B = c^2C$.

$a^2A + b^2B = c(cC) = c(aA + bB)$, or

$(A + B)\left(a^2A + b^2B\right) = (A + B)c(aA + bB) = cC(aA + bB)$, or

$a^2A + b^2AB + a^2AB + b^2B^2 = (aA + bB)^2 = a^2A^2 + 2abAB + b^2B^2$, or

$\left(a^2 + b^2 - 2ab\right)AB = 0$, or $(a - b)^2AB = 0$. But A and B are nonzero, so $a = b$. Therefore $(A + B)e^{iax} = Ce^{icx}$. $a(A + B) = cC$, or $aC = cC$, so (since $C \neq 0$) $a = c$. Conclusion: $a = b = c$ is driven.

Example 2: Analyze the case of polarization *perpendicular* to the plane of incidence (i.e., electric fields in the y direction, in Fig. F.22). Impose the boundary condition Eq. F.76, and obtain the Fresnel equations for $\widetilde{\vec{E}}_{0_R}$ and $\widetilde{\vec{E}}_{0_T}$. Sketch $(\widetilde{\vec{E}}_{0_R}/\widetilde{\vec{E}}_{0_I})$ and $(\widetilde{\vec{E}}_{0_T}/\widetilde{\vec{E}}_{0_I})$ as functions of , for the case $\beta = n_2/n_1 = 1.5$. (Note that for this β the reflected wave is always 180° out of phase). Show that there is no Brewster's angle for *any* n_1 and n_2: $\widetilde{\vec{E}}_{0_R}$ is *never* zero (unless, of course, $n_1 = n_2$ and $\mu_1 = \mu_2$, in which case the two media are optically indistinguishable). Confirm that your Fresnel equations reduce to the proper forms at normal incidence. Compute the reflection and transmission coefficients, and check that they add up to 1.

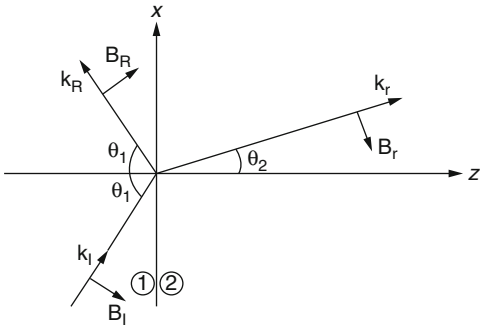
Solution: We start with the following relationships

F

$$\left\{\begin{array}{l}\widetilde{\vec{E}}_I = \widetilde{E}_{0_I} e^{i\left(\vec{k}_I \cdot \vec{r} - \omega t\right)} \hat{\vec{y}} \\ \widetilde{\vec{B}}_I = \dfrac{1}{v_1} \widetilde{E}_{0_I} e^{i\left(\vec{k}_I \cdot \vec{r} - \omega t\right)} \left(-\cos\theta_1 \hat{\vec{x}} + \sin\theta_1 \hat{\vec{z}} \right)\end{array}\right\}$$

$$\left\{\begin{array}{l}\widetilde{\vec{E}}_R = \widetilde{E}_{0_R} e^{i\left(\vec{k}_R \cdot \vec{r} - \omega t\right)} \hat{\vec{y}} \\ \widetilde{\vec{B}}_I = \dfrac{1}{v_1} \widetilde{E}_{0_R} e^{i\left(\vec{k}_R \cdot \vec{r} - \omega t\right)} \left(-\cos\theta_1 \hat{\vec{x}} + \sin\theta_1 \hat{\vec{z}} \right)\end{array}\right\}$$

$$\left\{\begin{array}{l}\widetilde{\vec{E}}_T = \widetilde{E}_{0_T} e^{i\left(\vec{k}_T \cdot \vec{r} - \omega t\right)} \hat{\vec{y}} \\ \widetilde{\vec{B}}_T = \dfrac{1}{v_1} \widetilde{E}_{0_R} e^{i\left(\vec{k}_T \cdot \vec{r} - \omega t\right)} \left(-\cos\theta_1 \hat{\vec{x}} + \sin\theta_1 \hat{\vec{z}} \right)\end{array}\right\}$$



Boundary conditions: $\left\{\begin{array}{lll}\text{(i)} & \varepsilon_1 E_1^{\perp} = \varepsilon_2 E_2^{\perp} & \text{(iii)} \quad E_1^{\|} = E_2^{\|} \\ \text{(ii)} & B_1^{\perp} = B_2^{\perp} & \text{(iv)} \quad \dfrac{1}{\mu_1} B_1^{\|} = \dfrac{1}{\mu_1} B_2^{\|}\end{array}\right.$

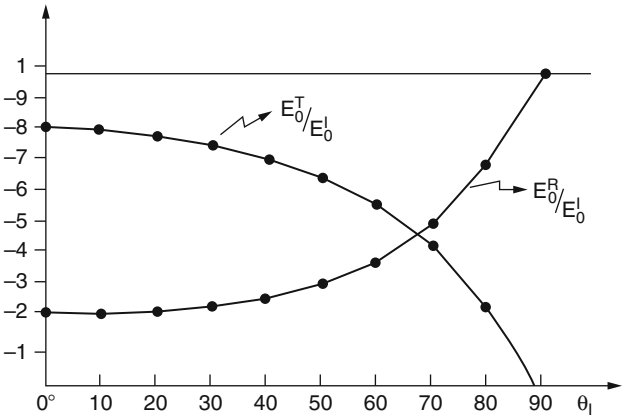
Law of refractions: $\dfrac{\sin\theta_2}{\sin\theta_1} = \dfrac{v_2}{v_1}$. [Note: $\vec{k}_I \cdot \vec{r} - \omega t = \vec{k}_R \cdot \vec{r} - \omega t = \vec{k}_T \cdot \vec{r} - \omega t$, at $z = 0$ so we can drop all exponential factors in applying the boundary conditions].

Boundary condition (i): 0=0 (trivial). Boundary condition (iii): $\widetilde{E}_{0_I} + \widetilde{E}_{0_R} = \widetilde{E}_{0_T}$.

Boundary condition (ii): $\dfrac{1}{v_1} \widetilde{E}_{0_I} \sin\theta_1 + \dfrac{1}{v_1} \widetilde{E}_{0_R} \sin\theta_1 = \dfrac{1}{v_2} \widetilde{E}_{0_r} \sin\theta_2 \Rightarrow$ $\widetilde{E}_{0_I} + \widetilde{E}_{0_R} = \left(\dfrac{v_1 \sin\theta_2}{v_2 \sin\theta_1} \right) \widetilde{E}_{0_r}$

But the term in parentheses is 1, by the law of refraction, so this is the same as (ii).

Boundary condition (iv): $\frac{1}{\mu_1}\left[\frac{1}{v_1}\widetilde{E}_{0_I}(-\cos\theta_1)+\frac{1}{v_1}\widetilde{E}_{0_x}(-\cos\theta_1)\right]=\frac{1}{\mu_2 v_2}\widetilde{E}_{0_r}(-\cos\theta_2)$ or we can write $\widetilde{E}_{0_I}-\widetilde{E}_{0_x}=\left(\frac{\mu_1 v_1\sin\theta_2}{\mu_2 v_2\sin\theta_1}\right)\widetilde{E}_{0_T}$. Let $\alpha\equiv\frac{\cos\theta_2}{\cos\theta_1}$ and $\beta\equiv\frac{\mu_1 v_1}{\mu_2 v_2}$. Then we have $\widetilde{E}_{0_I}-\widetilde{E}_{0_R}=\alpha\beta\widetilde{E}_{0_T}$. Solving for $\widetilde{E}_{0_R}$ and $\widetilde{E}_{0_T}$: $2\widetilde{E}_{0_I}=(1+\alpha\beta)\widetilde{E}_{0_T}$ or $\widetilde{E}_{0_T}=\left(\frac{2}{1+\alpha\beta}\right)\widetilde{E}_{0_I}$ and $\widetilde{E}_{0_R}=\widetilde{E}_{0_T}-\widetilde{E}_{0_I}=\left(\frac{2}{1+\alpha\beta}-\frac{1+\alpha\beta}{1+\alpha\beta}\right)\widetilde{E}_{0_I}$ then we have $\widetilde{E}_{0_T}=\left(\frac{1-\alpha\beta}{1\_\alpha\beta}\right)\widetilde{E}_{0_I}$. Since α and β are positive, it follows that $2/(1+\alpha\beta)$ is positive, and hence the *transmitted* wave is in *phase* with the incident wave, and the (real) amplitudes are related by $E_{0_T}=\left(\frac{2}{1+\alpha\beta}\right)E_{0_I}$. The *reflected* wave is in phase if $\alpha\beta<1$ and 180° out of phase if $\alpha\beta<1$; the (real) amplitudes are related by $E_{o_R}=\left(\frac{1-\alpha\beta}{1+\alpha\beta}\right)E_{0_I}$. These are the **Fresnel Equations** for polarization perpendicular to the plane of incidence. To construct the graphs, note that $\alpha\beta=\beta\frac{\sqrt{2.25-\sin^2\theta}}{\cos\theta}$, where θ is the angle of incidence, so, for $\beta=1.5$ and $\alpha\beta=\beta\frac{\sqrt{1-\sin^2\theta}}{\cos\theta}=\frac{\sqrt{\beta^2-\sin^2\theta}}{\cos\theta}$



Is there a Brewster's angle?. Well, $E_{0_R}=0$ would mean that $\alpha\beta=1$, and hence that $\alpha=\frac{\sqrt{1-(v_2/v_1)^2\sin^2\theta}}{\cos\theta}=\frac{1}{\beta}=\frac{\mu_2 v_2}{\mu_1 v_1}$ or $1-\left(\frac{v_2}{v_1}\right)^2\sin^2\theta=\left(\frac{\mu_2 v_2}{\mu_1 v_1}\right)\cos^2\theta$, so $1=\left(\frac{v_2}{v_1}\right)^2[\sin^2\theta+(\mu_2/\mu_1)\cos^2\theta]$. Since $\mu_1\approx\mu_2$, this means $1\approx(v_2/v_1)^2$, which is only true for optically indistinguishable media, in which case there is of

course no reflection—but that would be true at *any* angle, not just at a special "Brewster's angle". [If μ_2 were substantially different from μ_1, and the relative velocities were just right, it *would* be possible to get a Brewster's angle for this case, at.

$$\left(\frac{v_2}{v_1}\right)^2 = 1 - \cos^2\theta + \left(\frac{\mu_2}{\mu_1}\right)\cos^2\theta$$

$$\Rightarrow \cos^2\theta = \frac{(v_1/v_2)^2 - 1}{(\mu_2/\mu_1)^2 - 1} = \frac{(\mu_2\varepsilon_2/\mu_1\varepsilon_1) - 1}{(\mu_2/\mu_1{}^2) - 1}$$

$$= \frac{(\varepsilon_2/\varepsilon_1) - (\mu_1/\mu_2)}{(\mu_2/\mu_1) - (\mu_1/\mu_2)}$$

But the media would be very peculiar. By the same token, δ_R is either always 0, or always π, for a given interface—it does not switch over as you change θ, the way it does for polarization in the plane of incidence. In particular, if $\beta = 3/2$, then $\alpha\beta > 1$,for $\alpha\beta = \dfrac{\sqrt{1 - \sin^2\theta}}{\cos\theta} > 1$ if $2.25 - \sin^2\theta > \cos^2\theta$, or $2.25 > \sin^2\theta + \cos^2\theta = 1.0$. In general, for $\beta > 1$, $\alpha\beta > 1$, and hence $\delta_R = \pi$. For $\beta < 1$, $\alpha\beta < 1$, and $\delta_R = 0$. At *normal incidence*, $\alpha = 1$, so Fresnel's equations reduce to $E_{0_T} = \left(\dfrac{2}{1+\beta}\right)E_{0_I}$ and $E_{0_R} = \left(\dfrac{1-\beta}{1+\beta}\right)E_{0_I}$ consistent with Eq. F.57. *Reflection* and *Transmission Coefficients*: $R = \left(\dfrac{E_{0_R}}{E_{0_I}}\right)^2 = \left(\dfrac{1-\alpha\beta}{1+\alpha\beta}\right)^2$. Referring to Eq. F.91, $T = \dfrac{\varepsilon_2 v_2}{\varepsilon_1 v_1}\alpha\left(\dfrac{E_{0_T}}{E_{0_I}}\right) = \alpha\beta\left(\dfrac{2}{\alpha+\beta}\right)^2$, therefore

$$R + T = \frac{(1-\alpha\beta)^2 + 4\alpha\beta}{(1+\alpha\beta)^2} = \frac{1 - 2\alpha\beta + \alpha^2\beta^2 + 4\alpha\beta}{(1+\alpha\beta)^2} = \frac{(1+\alpha\beta)^2}{(1+\alpha\beta)^2} = 1$$

Example 3: The index of refraction of diamond is 2.42. Construct the graph analogous to Fig. F.23 for the air/diamond interface. (Assume $\mu_1 = \mu_2 = \mu_0$) In particular, calculate (a) the amplitudes at normal incidence, (b) Brewster's angle, and (c) the "crossover" angle, at which the reflected and transmitted amplitudes are equal.

Solution: Equation F.81 we see that $\beta = 2.42$, Equation F.85 $\alpha = \dfrac{\sqrt{1 - (\sin\theta/2.42)^2}}{\cos\theta}$

$\theta = 0 \Rightarrow \alpha = 1$. Equation 9.109 $\Rightarrow \left(\dfrac{E_{0_R}}{E_{0_I}}\right) = \dfrac{\alpha-\beta}{\alpha+\beta} = \dfrac{1-2.42}{1+2.42} = -\dfrac{1.42}{3.42} = -0.415$

$$\left(\frac{E_{0_T}}{E_{0_I}}\right) = \frac{2}{\alpha+\beta} = \frac{2}{3.42} = 0.585$$

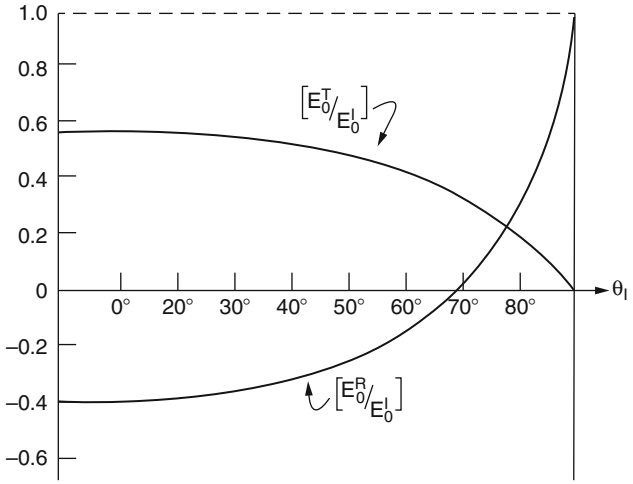
Equation F.87 $\Rightarrow \theta_B tn^{-1}(2.42) = 67.5°$

$E_{0_R} = E_{0_T} \Rightarrow \alpha - \beta = 2 \quad ; \qquad \alpha = \beta + 2 = 4.42 \qquad \text{and} \qquad (4.42)^2 \cos^2\theta = 1 - \sin^2\theta/(2.42)^2$

$$(4.42)^2\left(1 - \sin^2\theta\right) = (4.42)^2 - (4.42)^2 \sin^2\theta = 1 - 0.171 \sin^2\theta$$

$$19.5 - 1 = (19.5 - 017) \sin^2\theta$$

$$18.5 = 19.3 \sin^2\theta; \ \sin^2\theta = 18.5/19.3 = 0.959 \Rightarrow \sin\theta = 0.979; \ \theta = 78.3^\circ$$



F.8 Absorption and Dispersion

Expressions for the stored energy, energy flow, and power dissipated are derived for electromagnetic waves in terms of the complex permittivity $\widetilde{\varepsilon}$ and permeability $\widetilde{\mu}$ for a frequency-dispersive absorbing medium. This is shown to be possible when $\widetilde{\varepsilon}$ and $\widetilde{\mu}$ are known functions not only of frequency but also of all the loss factors—e.g., collision frequencies, etc. The derivation is not restricted to media with small losses.

In physics and electrical engineering, dispersion most often refers to frequency-dependent effects in wave propagation. Note, however, that there are several other uses of the word "dispersion" in the physical sciences. In the presence of dispersion, wave velocity is no longer uniquely defined, giving rise to the distinction of phase velocity and group velocity. A well-known effect of phase velocity dispersion is the color dependence of light refraction that can be observed in prisms and rainbows.

Dispersion relations describe the interrelations of wave properties like wavelength, frequency, velocities, refraction index, and attenuation coefficient. Besides geometry- and material-dependent dispersion relations, there are the overarching Kramers–Kronig relations that connect the frequency dependences of propagation and attenuation.

Dispersion may be caused either by geometric boundary conditions (waveguides, shallow water) or by interaction of the waves with the transmitting medium.

F.8.1 Electromagnetic Waves in Conductors

Previous sections were stipulating that the free charge density ρ_f and the free current density $\vec{J}_f$ are zero, and everything that followed was predicated on that assumption. Such a restriction is perfectly reasonable when you're talking about wave propagation through a vacuum or through insulating materials such as glass or (pure) water. But in the case of conductors we do not independently control the flow of charge, and in general $\vec{J}_f$ is certainly not zero. In fact, according to Ohm's law, the (free) current density in a conductor is proportional to the electric field:

$$\vec{J}_f = \sigma \vec{E} \tag{F.92}$$

With this, Maxwell's equations for linear media assume the form

$$\left.\begin{array}{ll} \text{(i)} \quad \nabla \cdot \vec{E} = \dfrac{1}{\varepsilon}\rho_f & \text{(iii)} \quad \nabla \times \vec{E} = \dfrac{\partial \vec{B}}{\partial t} \\ \text{(ii)} \quad \nabla \cdot \vec{B} = 0 & \text{(iv)} \quad \nabla \times \vec{B} = \mu\sigma\vec{E} + \mu\varepsilon \dfrac{\partial \vec{E}}{\partial t} \end{array}\right\} \tag{F.93}$$

Now the continuity equation for free charge,

$$\nabla \cdot \vec{J}_f = -\frac{\partial \rho_f}{\partial t} \tag{F.94}$$

together with Ohm's law and Gauss's law (i), gives

$$\frac{\partial \rho_f}{\partial t} = -\sigma\left(\nabla \cdot \vec{E}\right) = -\frac{\sigma}{\varepsilon}\rho_f \tag{F.95}$$

for a homogeneous linear medium, from which it follows that

$$\rho_f(t) = e^{-(\sigma/\varepsilon)t}\rho_f(0) \tag{F.96}$$

Thus any initial free charge density $\rho_f(0)$ dissipates in a characteristic time $\tau \equiv \varepsilon/\sigma$. This reflects the familiar fact that if you put some free charge on a conductor, it will flow out to the edges. The time constant τ affords a measure of how "good" a conductor is: For a "perfect" conductor $\sigma = \infty$ and $\tau = 0$; for a "good" conductor, τ is much less than the other relevant times in the problem (in oscillatory systems, that means $\tau \ll 1/\omega$; for a "poor" conductor, τ is *greater* than the characteristic times in the problem ($\tau \gg 1/\omega$) [5]. N. Ashby [5] points out that for good conductor conductors, is absurdly short (10^{-19}s, for copper, whereas the time between collisions is $\tau_c = 10^{-14}$s). The problem is that Ohm's law itself breaks down on time scales shorter than τ_c; actually, the time it takes free charge to dissipate in a

good conductor is of order τ_c, not T. Moreover [6], shows that it takes even longer for the fields and currents to equilibrate. But none of this is relevant to our present purpose; the free charge density in a conductor does eventually dissipate, and exactly how long the process takes is beside the point.

At present we're not interested in this transient behavior-we'll wait for any accumulated free charge to disappear. From then on $\rho_f = 0$, and we have

$$\left.\begin{array}{ll} \text{(i)} & \nabla \cdot \vec{E} = 0 \quad \text{(iii)} \quad \nabla \times \vec{E} = \dfrac{\partial \vec{B}}{\partial t} \\ \text{(ii)} & \nabla \cdot \vec{B} = 0 \quad \text{(iv)} \quad \nabla \times \vec{B} = \mu\sigma\vec{E} + \mu\varepsilon\dfrac{\partial \vec{E}}{\partial t} \end{array}\right\} \tag{F.97}$$

These differ from the corresponding equations for non-conducting media (Eq. F.46) only in the addition of the last term in (iv). Applying the curl to (iii) and (iv), as before, we obtain modified wave equations for $\vec{E}$ and $\vec{B}$:

$$\nabla \vec{E} = \mu\varepsilon\frac{\partial^2 \vec{E}}{\partial t^2} + \mu\sigma\frac{\partial \vec{E}}{\partial t} \quad \text{and} \quad \nabla \vec{B} = \mu\varepsilon\frac{\partial^2 \vec{B}}{\partial t^2} + \mu\sigma\frac{\partial \vec{B}}{\partial t} \tag{F.98}$$

These equations still admit plane-wave solutions,

$$\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{i\left(\widetilde{k}z - \omega t\right)} \quad \text{and} \quad \widetilde{\vec{B}}(z,t) = \widetilde{\vec{B}}_0 e^{i\left(\widetilde{k}z - \omega t\right)} \tag{F.99}$$

but this time the "wave number" $\widetilde{k}$ is complex:

$$\widetilde{k}_2 = \mu\varepsilon\omega^2 + i\mu\sigma\omega \tag{F.100}$$

as you can easily check by plugging Eq. F.99 into Eq. F.98. Taking the square root,

$$\widetilde{k} = k + i\kappa \tag{F.101}$$

where

$$k \equiv \omega\sqrt{\frac{\varepsilon\mu}{2}\left[\sqrt{1 + \left(\frac{\sigma}{\varepsilon\omega}\right)^2} + 1\right]}^{1/2} \quad \textit{and} \quad \kappa \equiv \omega\sqrt{\frac{\varepsilon\mu}{2}\left[\sqrt{1 + \left(\frac{\sigma}{\varepsilon\omega}\right)^2} - 1\right]}^{1/2} \tag{F.102}$$

The imaginary part of k results in an attenuation of the wave (decreasing amplitude with increasing z):

$$\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{-\kappa z} e^{i(kz - \omega t)} \quad \text{and} \quad \widetilde{\vec{B}}(z,t) = \widetilde{\vec{B}}_0 e^{-\kappa z} e^{i(kz - \omega t)} \tag{F.103}$$

The distance it takes to reduce the amplitude by a factor of $1/e$ (about a third) is called the **skin depth**:

$$d \equiv \frac{1}{\kappa} \tag{F.104}$$

it is a measure of how far the wave penetrates into the conductor [1]. Meanwhile, the real part of $\widetilde{k}$ determines the wavelength, the propagation speed, and the index of refraction, in the usual way:

$$\lambda = \frac{2\pi}{k}, \quad v = \frac{\omega}{k}, \quad n = \frac{ck}{\omega} \tag{F.105}$$

The attenuated plane waves (Eq. F.103) satisfy the modified wave Eq. (F.98) for any $\widetilde{\vec{E}}_0$ and $\widetilde{\vec{B}}_0$. But Maxwell's equations (F.97) impose further constraints, which serve to determine the relative amplitudes, phases, and polarizations of $\vec{E}$ and $\vec{B}$. As before, (i) and (ii) rule out any z components: the fields are transverse. We may as well orient our axes so that $\vec{E}$ is polarized along the x-direction:

$$\widetilde{\vec{E}}(z,t) = \widetilde{E}_0 e^{-\kappa z} e^{i(kz-\omega t)}\,\hat{x} \tag{F.106}$$

Then (iii) gives

$$\widetilde{\vec{B}}(z,t) = \frac{\widetilde{k}}{\omega}\widetilde{E}_0 e^{-\kappa z} e^{i(kz-\omega t)}\,\hat{y} \tag{F.107}$$

Equation (iv) says the same thing). Once again, the electric and magnetic fields are mutually perpendicular. Like any complex number, $\widetilde{k}$ can be expressed in terms of its modulus and phase [1]:

$$k = Ke^{i\phi} \tag{F.108}$$

where

$$K = |\widetilde{k}| = \sqrt{k^2 + \kappa^2} = \omega\sqrt{\varepsilon\mu}\sqrt{1 + \left(\frac{\sigma}{\varepsilon\omega}\right)^2} \tag{F.109}$$

and

$$\phi \equiv \tan^{-1}(\kappa/k)$$

According to Eqs. F.104 and F.107, the complex amplitudes $\widetilde{E}_0 = E_0 e^{i\delta_E}$ and $\widetilde{B}_0 = B_0 e^{i\delta_B}$ are related by

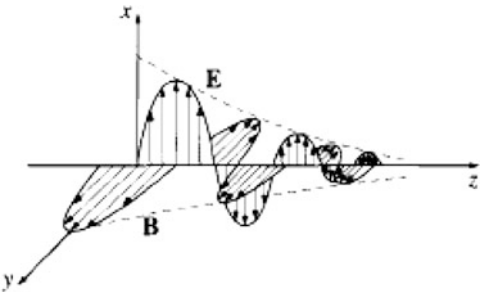


Fig. F.26 Presentation of electric and magnetic fields

$$B_0 e^{i\delta_B} = \frac{K e^{i\phi}}{\omega} e^{i\delta_B} \tag{F.110}$$

Evidently the electric and magnetic fields are no longer in phase; in fact,

$$\delta_B - \delta_E = \phi \tag{F.111}$$

The magnetic field *lags behind* the electric filed [1]. Meanwhile, the (real) amplitudes of $\vec{E}$ and $\vec{B}$ are related by

$$\frac{B_0}{E_0} = \frac{K}{\omega}\sqrt{\varepsilon\mu\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2}} \tag{F.112}$$

The (real) electric and magnetic filds are, finally,

$$\left.\begin{aligned}\vec{E}(z,t) &= E_0 e^{-\kappa z}\cos(kz-\omega t+\delta_E)\hat{\vec{x}}\\ \vec{B}(z,t) &= B_0 e^{-\kappa z}\cos(kz-\omega t+\delta_E+\phi)\hat{\vec{y}}\end{aligned}\right\} \tag{F.113}$$

These fields are shown in Fig. F.26 below.

Example 1: Suppose you imbedded some free charge in a piece of glass. About how long would it take for the charge to flow to the surface?

Silver is an excellent conductor, but it's expensive. Suppose you were designing a microwave experiment to operate at a frequency of 1010 Hz. How thick would you make the silver coatings?

Find the wavelength and propagation speed in copper for radio waves at 1 MHz. Compare the corresponding values in air (or vacuum).

Solution: Equation F.96 $\Rightarrow \tau = \varepsilon/\sigma$. But $\varepsilon = \varepsilon_0\varepsilon_r$ and from Eq. F.49, we have $\varepsilon_r \cong n^2$, and for glass the index of refraction is typically around 1.5, so $\varepsilon \simeq (1.5)^2 kt 8.85 \times 10^{-12} C^2/Nm^2$, while $\sigma = 1/\rho \simeq 10^{-12}\Omega m$ (Table F.3 below). Then $\tau = \left(2\times 10^{-11}\right)/10^{-12} = 20\,\text{sec}$. (But the resistivity of glass varies enormously from one type to another, so this answer could be off by a factor of 100 in either direction).

Table F.3 Resistivities, in ohm-meters (all values are for 1 atm, 200 °C) [7]

Material	Resistivity	Material	Resistivity
Conductors:		Semiconductors:	
Silver	1.59×10^{-8}	Salt water (saturated)	4.4×10^{-2}
Copper	1.68×10^{-8}	Germanium	4.6×10^{-1}
Gold	2.21×10^{-8}	Diamond	2.7
Aluminum	2.65×10^{-8}	Silicon	2.5×10^{3}
Iron	9.61×10^{-7}	Insulators:	
Mercury	9.58×10^{-7}	Water (pure)	2.5×10^{5}
Nichrome	1.00×10^{-6}	Wood	$10^{8} - 10^{11}$
Manganese	1.44×10^{-6}	Glass	$10^{10} - 10^{14}$
Graphite	1.4×10^{-5}	Quartz (fused)	$\sim 10^{16}$

For silver, $\rho = 1.59 \times 10^{-8}$ (Table F.3), and $\varepsilon \approx \varepsilon_0$, so $\omega\varepsilon = 2\pi \times 10^{10} \times 8.85 \times 10^{-12} = 0.56$. Since $\sigma = 1/\rho = 6.25 \times 10^{7} \Omega m \gg \omega\varepsilon$, the skin depth (Eq. F.104) is $d = \frac{1}{\kappa} \cong \sqrt{\frac{2}{\omega\sigma\mu}} = \sqrt{\frac{2}{2\pi \times 10^{10} \times 6.25 \times 10^{7} \times 4\pi \times 10^{-7}}} = 6.4 \times 10^{-77} m = 6.4 \times 10^{-4} mm$ we place plate silver to a depth of about 0.001 mm; there is no point in making it any thicker, since the fields do not penetrate much beyond this anyway.

For copper, Table F.3 above gives $\sigma = 1/\left(1.68 \times 10^{-8}\right) \approx 6 \times 10^{7}$, $\omega\varepsilon_0 = \left(2\pi \times 10^{6}\right) \times \left(8.85 \times 10^{-12}\right) = 8.85 \times 10^{-5}$. Since $\sigma \gg \omega\varepsilon$, Eq. F.103 $\Rightarrow k \approx 2\pi\sqrt{\frac{2}{\omega\sigma\mu_0}} = 2\pi\sqrt{\frac{2}{2\pi \times 10^{6} \times 6 \times 10^{7} \times 4\pi \times 10^{-7}}} = 4 \times 10^{-7} m = 0.4mm$

From Eq. F.103, the propagation speed is $v = \frac{\omega}{k} = \frac{\omega}{2\pi}\lambda = \lambda\nu = \left(4 \times 10^{-4}\right) \times 10^{6} = 400 m/s$. In vacuum, $\lambda = \frac{c}{\nu} = \frac{4 \times 10^{8}}{10^{6}} = 300m$; $v = c = 44 \times 10^{8} m/s$. (But really, in a good conductor the skin depth is so small, compared to the wavelength, that the notions of "wavelength" and "propagation speed" lose their meaning.

F.8.2 Reflection at a Conducting Surface

The boundary conditions we used to analyze reflection and refraction at an interface between two dielectrics do not hold in the presence of free charges and currents. Instead, we have the more general relations (Eq. F.42) with addition term of surface current $\vec{k}_f$ in (iv) [1]:

$$\left.\begin{array}{ll} \text{(i)} \quad \varepsilon_1 E_1^{\perp} - \varepsilon_2 E_2^{\perp} = 0 & \text{(iii)} \quad E_1^{\parallel} - E_2^{\parallel} = 0 \\ \text{(ii)} \quad B_1^{\perp} - B_2^{\perp} = 0 & \text{(iv)} \quad \dfrac{1}{\mu_1} B_1^{\parallel} \dfrac{1}{\mu_2} B_2^{\parallel} = \vec{K}_f \times \hat{n} \end{array}\right\} \tag{F.114}$$

In this case σ_f (not to be confused with conductivity) is the free surface charge $\vec{k}_f$ the free surface current, and $\hat{n}$ (not to be confused with the polarization of the wave) is a unit vector perpendicular to the surface, pointing from medium (F.2) into medium (F.1). For ohmic conductors ($\vec{J}_f = \sigma \vec{E}$) there can be no free surface current, since this would require an infinite electric field at the boundary [1].

Suppose now that the xy-plane forms the boundary between a non-conducting linear medium (F.1) and a conductor (F.2). A monochromatic plane wave, traveling in the z-direction and polarized in the x direction, approaches from the left, as in Fig. F.20:

$$\widetilde{\vec{E}}_I(z,t) = \widetilde{E}_{0_I} e^{i(k_1 z - \omega t)} \hat{x} \quad \text{and} \quad \widetilde{\vec{B}}_I(z,t) = \frac{1}{v_1} \widetilde{B}_{0_I} e^{i(k_1 z - \omega t)} \hat{y} \tag{F.115}$$

This incident wave gives rise to a reflected wave [1]

$$\widetilde{\vec{E}}_R(z,t) = \widetilde{E}_{0_R} e^{i(-k_1 z - \omega t)} \hat{x} \quad \text{and} \quad \widetilde{\vec{B}}_R(z,t) = \frac{1}{v_1} \widetilde{B}_{0_R} e^{i(-k_1 z - \omega t)} \hat{y} \tag{F.116}$$

Propagating back to the left in medium (F.1), and a transmitted wave

$$\widetilde{\vec{E}}_T(z,t) = \widetilde{E}_{0_T} e^{i\left(\widetilde{k}_2 z - \omega t\right)} \hat{x} \quad \text{and} \quad \widetilde{\vec{B}}_T(z,t) = \frac{1}{v_1} \widetilde{B}_{0_T} e^{i\left(\widetilde{k}_2 z - \omega t\right)} \hat{y} \tag{F.117}$$

which is attenuated as it penetrates into the conductor.

At $z = 0$, the combined wave in medium (F.1) must join the wave in medium (F.2), pursuant to the boundary conditions Eq. F.114. Since $E^{\perp} = 0$ on both sides, boundary condition (i) yields $\sigma_f = 0$. Since $B^{\perp} = 0$, (ii) is automatically satisfied. Meanwhile, (iii) gives;

$$\widetilde{E}_{0_I} + \widetilde{E}_{0_R} = \widetilde{E}_{0_T} \tag{F.118}$$

and (iv) (with $K_f = 0$) says

$$\frac{1}{\mu_1 v_1} \left(\widetilde{E}_{0_I} - \widetilde{E}_{0_R} \right) - \frac{\widetilde{k}_2}{\mu_2 \omega} \widetilde{E}_{0_T} = 0 \tag{F.119}$$

or

$$\widetilde{E}_{0_I} - \widetilde{E}_{0_R} = \widetilde{\beta}\widetilde{E}_{0_T} \tag{F.120}$$

where

$$\widetilde{\beta} \equiv \frac{\mu_1 v_1}{\mu_2 \omega}\widetilde{k}_2 \tag{F.121}$$

It follows that

$$\widetilde{E}_{0_R} = \left(\frac{1-\widetilde{\beta}}{1+\beta}\right)\widetilde{E}_{0_I} \quad \text{and} \quad \widetilde{E}_{0_T} = \left(\frac{2}{1+\beta}\right)\widetilde{E}_{0_I} \tag{F.122}$$

These results are formally identical to the ones that apply at the boundary between *nonconductors* (Eq. F.57), but the resemblance is deceptive since $\widetilde{\beta}$ is now a complex number.

For a *perfect* conductor ($\sigma = \infty$), $k_2 = \infty$ (Eq. F.102), so $\widetilde{\beta}$ is infinite, and

$$\widetilde{E}_{0_R} = -\widetilde{E}_{0_I} \quad \text{and} \quad \widetilde{E}_{0_T} = 0 \tag{F.123}$$

In this case the wave is totally reflected, with a 180° phase shift. (That's why excellent conductors make good mirrors. In practice, you paint a thin coating of silver onto the back of a pane of glass—the glass has nothing to do with the reflection; it's just there to support the silver and to keep it from tarnishing. Since the skin depth in silver at optical frequencies is on the order of 100 Å, you don't need a very thick layer).

Example 1: Calculate the reflection coefficient for light at an air-to-silver interface ($\mu_1 = \mu_2 = \mu_0$, $\varepsilon_1 = \varepsilon_0$, $\sigma = 6 \times 10^7 (\Omega \cdot m)^{-1}$, at optical frequencies ($\omega = 4 \times 10^{15}\ s^{-1}$)

Solution: According to Eq. F.122, $R = \left|\frac{\widetilde{E}_{0_R}}{\widetilde{E}_{0_I}}\right|^2 = \left|\frac{1-\widetilde{\beta}}{1+\beta}\right|^2 = \left(\frac{1-\widetilde{\beta}}{1-\beta}\right)\left(\frac{1-\widetilde{\beta}^*}{1+\beta^*}\right)$, where $\widetilde{\beta}^*$ is complex conjugate of $\widetilde{\beta}$ and $\widetilde{\beta} = \frac{\mu_1 v_1}{\mu_2 \omega}\widetilde{k}_2 = \frac{\mu_1 v_1}{\mu_2 \omega}(k_2 + i\kappa_2)$ (Eqs. F.101 and F.121). Since silver is a good conductor ($\sigma \gg \varepsilon\omega$), Eq. F.102 reduces to $\kappa_2 \cong k_2 \cong \omega\sqrt{\frac{\varepsilon_2\mu_2}{2}}\sqrt{\frac{\sigma}{\varepsilon_2\omega}} = \sqrt{\frac{\sigma\omega\mu_2}{2}}$, so $\widetilde{B} = \frac{\mu_1 v_1}{\mu_2 \omega}\sqrt{\frac{\sigma\omega\mu_2}{2}}(1+i) = \mu_1 v_1\sqrt{\frac{\sigma}{2\mu_2\omega}}(1+i)$. Now let $\gamma \equiv \mu_1 v_1\sqrt{\frac{\sigma}{2\mu_2\omega}} = \mu_0 c\sqrt{\frac{\sigma}{2\mu_0\omega}} = c\sqrt{\frac{\sigma\mu_0}{2\omega}} = \left(3\times 10^8\right)\sqrt{\frac{\left(3\times 10^7\right)\left(4\pi\times 10^{-7}\right)}{(2)\left(4\times 10^{15}\right)}} = 29$ $R = \left(\frac{1-\gamma-i\gamma}{1+\gamma+i\gamma}\right)\left(\frac{1-\gamma+i\gamma}{1+\gamma-i\gamma}\right) = \frac{(1-\gamma)^2+\gamma^2}{(1+\lambda)^2+\gamma^2} = 0.93$. Evidently 93 % of the light is reflected.

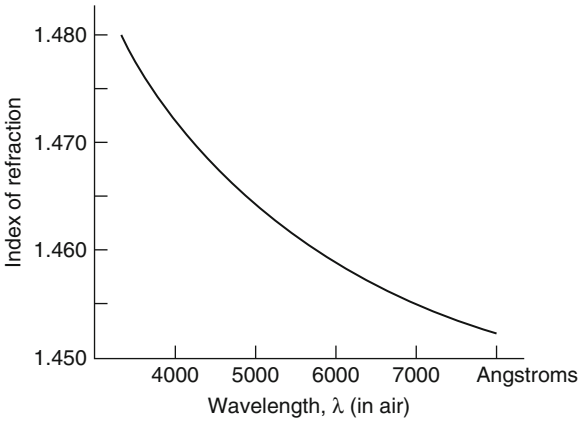


Fig. F.27 Graph for typical class

F.8.3 The Frequency Dependence of Permittivity

In the preceding sections, we have seen that the propagation of electromagnetic waves through matter is governed by three properties of the material, which we took to be constants and they are as follows:

1. The permittivity ε
2. The permeability μ
3. The conductivity σ

Actually, each of these parameters depends to some extent on the frequency of the waves you are considering. Indeed, if the permittivity were *truly* constant, then the index of refraction in a transparent medium, $n \cong \sqrt{\varepsilon_r}$, would also be constant. But it is well known from optics that n is a function of wavelength (Fig. F.27 shows the graph for a typical glass). A prism or a raindrop bends blue light more sharply than red, and spreads white light out into a rainbow of colors. This phenomenon is called **dispersion**. By extension, whenever the speed of a wave depends on its frequency, the supporting medium is called **dispersive**. Conductors, incidentally, are dispersive, see Eqs. F.102 and F.103.

Because waves of different frequency travel at different speeds in a dispersive medium, a wave form that incorporates a range of frequencies will change shape as it propagates. A sharply peaked wave typically flattens out, and whereas each sinusoidal component travels at the ordinary **wave** (or **phase**) velocity,

$$v = \frac{\omega}{k} \tag{F.124}$$

the packet as a whole (the "envelope") travels at the so-called **group velocity** [8]. Also for more information refer to Appendix G of this book.

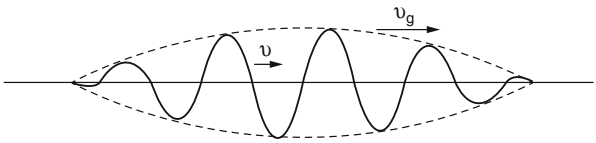


Fig. F.28 Phase and group velocity [1]

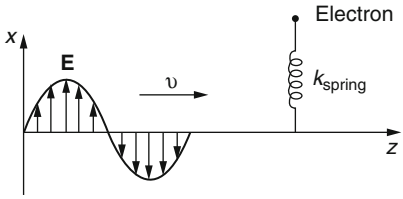


Fig. F.29 Electron movement [1]

$$\upsilon_g = \frac{d\omega}{dk} \tag{F.125}$$

Figure F.28 is a typical depict of these two wave velocities.

Our purpose in this section is to account for the frequency dependence of ε in nonconductors, using a simplified model for the behavior of electrons in dielectrics. The electrons in nonconductor are bound to specific molecules [1]. The actual binding forces can be quite complicated, but we shall picture each electron as attached to the end of an imaginary spring, with force constant k_{spring} (Fig. F.29).

$$F_{binding} = k_{spring}x = -m\omega_0^2 x \tag{F.126}$$

where x is displacement from equilibrium, m is the electron's mass, and ω_0 is the natural oscillation frequency, $\sqrt{k_{spring}/m}$. Utilizing Taylor series and expanding the potential energy about equilibrium point for sufficiently small displacement we have [1]:

$$U(x) = U(0) + xU^{'}(0) + \frac{1}{2}x^2U^{''}(0) + \cdots$$

The first term is a constant, with no dynamical significance (you can always adjust the zero of potential energy so that $U(0) = 0$). The second term automatically vanishes, since $dU/dx = -F$, and by the nature of an equilibrium the force at that point is zero. The third term is precisely the potential energy of a spring with force constant $k_{spring} = d^2U/dx^2\big|_0$ (the second derivative is positive, for a point of stable equilibrium). As long as the displacements are small, the higher terms in the series can be neglected [1]. Meanwhile, there will presumably be some damping force on the electron:

$$F_{damping} = -my\frac{dx}{dt} \tag{F.127}$$

The damping must be opposite in direction to the velocity, and making it *proportional* to the velocity is the easiest way to accomplish this. The *cause* often damping does not concern us here—among other things, an oscillating charge radiates, and the radiation siphons off energy.

In the presence of an electromagnetic wave of frequency ω, polarized in the x (Fig. F.29) direction, the electron is subject to a driving force

$$F_{driving} = qE = qE_o \cos(\omega t) \tag{F.128}$$

where q is the charge of the electron and E_0 is the amplitude of the wave at the point z where the electron is situated. If we concentrate only at one point and one point only so that the maximum E occurs there at $t = 0$, then by Newton's second law we have:

$$\begin{aligned} m\frac{d^2x}{dt^2} &= F_{total} = F_{binding} + F_{damping} + F_{driving} \\ m\frac{d^2x}{dt^2} &+ my\frac{dx}{dt} + m\omega_0^2 x = qE_0 \cos(\omega t) \end{aligned} \tag{F.129}$$

Our model, then, describes the electron as a damped harmonic oscillator, driven at frequency ω. (assumption is that the much more massive nuclei remain at rest) [1].

Equation F.129 is easier to handle if we regard it as the real part of a *complex equation*:

$$\frac{d^2\tilde{x}}{dt^2} + \gamma\frac{d\tilde{x}}{dt} + \omega_0^2\tilde{x} = \frac{q}{m}E_0 e^{-i\omega t} \tag{F.130}$$

In steady state, the system oscillates at the driving frequency:

$$\tilde{x}(t) = \tilde{x}_0 e^{-i\omega t} \tag{F.131}$$

Inserting this into Eq. F.130, we obtain

$$\tilde{x}_0 = \frac{q/m}{\omega_0^2 - \omega^2 - i\gamma\omega}E_0 \tag{F.132}$$

The dipole moment is the real part of

$$\widetilde{p}(t) = q\widetilde{x}(t) = \frac{q/m}{\omega_0^2 - \omega^2 - i\gamma\omega} E_0 e^{-i\omega t} \tag{F.133}$$

The imaginary term in the denominator means that p is *out of phase* with E—lagging behind by angle $\tan^{-1}\left[\gamma\omega/\left(\omega_0^2 - \omega^2\right)\right]$ that is very small when $\omega \ll \omega_0$ and rises to π when $\omega \gg \omega_0$.

In general, differently situated electrons within a given molecule experience different natural frequencies and damping coefficients. Let's say there are f_j electrons with frequency ω_j and damping γ_j in each molecule. If there are N molecules per unit volume, the polarization $\mathbf{P}$ is given by the real part of

$$\widetilde{\vec{\mathbf{P}}} == \frac{Nq^2}{m}\left(\sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\gamma_j\omega}\right)\widetilde{\vec{\mathbf{E}}} \tag{F.134}$$

This applies directly to the case of a dilute gas; for denser materials the theory is modified slightly, in accordance with the Clausius–Mossotti Equation. Note that we should not confuse the "polarization" of a medium, $\vec{\mathbf{P}}$, with the "polarization" of a *wave—same word*, but two completely unrelated meanings.

Now we define the electric susceptibility as the proportionality constant between $\vec{\mathbf{P}}$ and $\widetilde{\vec{\mathbf{E}}}$ (specifically, $\vec{\mathbf{P}} = \varepsilon_0\chi_e\vec{\mathbf{E}}$) [1]. In the present case $\vec{\mathbf{P}}$ is *not* proportional to $\widetilde{\vec{\mathbf{E}}}$ (this is not, strictly speaking, a linear medium) because of the difference in phase. However, the *complex* polarization $\vec{\mathbf{P}}$ is proportional to the *complex* field $\widetilde{\vec{\mathbf{E}}}$, and this suggests that we introduce a **complex susceptibility**, $\widetilde{\chi}_e$:

$$\vec{\mathbf{P}} = \varepsilon_0\widetilde{\chi}_e\vec{\mathbf{E}} \tag{F.135}$$

From all these we conclude that the physical polarization is the real part of $\widetilde{\vec{\mathbf{P}}}$, just as the physical field is the real part of $\widetilde{\vec{\mathbf{E}}}$. In particular, the proportionality between $\widetilde{\vec{\mathbf{D}}}$ and $\widetilde{\vec{\mathbf{E}}}$ is the **complex permittivity** $\widetilde{\varepsilon} = \varepsilon_0(1 + \widetilde{\chi}_e)$, and the **complex dielectric constant** in this model is

$$\widetilde{\varepsilon}_r = 1 + \frac{Nq^2}{m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\gamma_j\omega} \tag{F.136}$$

Ordinarily, the imaginary term is negligible; however, when ω is very close to one of the resonant frequencies (ω_j) it plays an important role, as we shall see. In a dispersive medium the wave equation for a given frequency reads

$$\frac{\nabla^2 \widetilde{\vec{\mathbf{E}}} = \widetilde{\varepsilon}\mu_0 \partial^2 \widetilde{\vec{\mathbf{E}}}}{\partial t^2} \tag{F.137}$$

This provides plane wave solutions, as before;

$$\widetilde{\vec{\mathbf{E}}}(z,t) = \widetilde{\vec{\mathbf{E}}}_0 e^{i\left(\widetilde{k}z - \omega t\right)} \tag{F.138}$$

with the complex wave number

$$\widetilde{k} \equiv \sqrt{\widetilde{\varepsilon}\mu_0 \omega} \tag{F.139}$$

Writing $\widetilde{k}$ in terms of its real and imaginary parts,

$$\widetilde{k} = k + i\kappa \tag{F.140}$$

Equation F.138 becomes

$$\widetilde{\vec{\mathbf{E}}}(z,t) = \widetilde{E}_0 e^{-\kappa z} e^{i(kz - \omega t)} \tag{F.141}$$

Evidently the wave is *attenuated* which is not surprising, since the damping absorbs energy [1]. Because the intensity is proportional to E^2 and hence to ${}_0e^{-2\kappa z}$, the quantity

$$\alpha \equiv 2\kappa \tag{F.142}$$

is called the **absorption coefficient**. Meanwhile, the wave velocity is ω/k, and the index of refraction is

$$n = \frac{ck}{\omega} \tag{F.143}$$

However, in the present case k and κ have nothing to do with conductivity; rather, they are determined by the parameters of our damped harmonic oscillator. For gases, the second term in Eq. F.136 is small, and we can approximate the square root (Eq. F.139) by the first term in the binomial expansion, $\sqrt{1+\varepsilon} \cong 1 + \frac{1}{2}\varepsilon$. Then

$$\widetilde{k} = \frac{\omega}{c}\sqrt{\widetilde{\varepsilon}_r} \cong \frac{\omega}{c}\left[1 + \frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\gamma_j\omega}\right] \tag{F.144}$$

or

$$n = \frac{ck}{\omega} \cong 1 + \frac{Nq^2}{m\varepsilon_0} \sum_j \frac{f_j\left(\omega_j^2 - \omega^2\right)}{\left(\omega_j^2 - \omega^2\right)^2 - i\gamma_j^2\omega^2} \tag{F.145}$$

and

$$\alpha = 2\kappa \cong \frac{Nq^2\omega^2}{m\varepsilon_0 c} \sum_j \frac{f_j\gamma_j}{\left(\omega_j^2 - \omega^2\right)^2 - i\gamma_j^2\omega^2} \tag{F.146}$$

Griffiths [1] has plotted the index of refraction and the absorption coefficient in the vicinity of one of the resonances. *Most* of the time the index of refraction *rises* gradually with increasing frequency, consistent with our experience from optics Fig. F.27. However, in the immediate neighborhood of a resonance the index of refraction drops sharply. Because this behavior is atypical, it is called **anomalous dispersion**. Notice that the region of anomalous dispersion ($\omega_1 < \omega < \omega_2$, in the figure) coincides with the region of maximum absorption; in fact, the material may be practically opaque in this frequency range. The reason is that we are now driving the electrons at their "favorite" frequency; the amplitude of their oscillation is relatively large, and a correspondingly large amount of energy is dissipated by the damping mechanism.

In Fig. F.30, n runs below 1above the resonance, suggesting that the wave speed exceeds c. This is no cause for alarm, since energy does not travel at the wave velocity but rather at the *group velocity* (see Example 1 below)). Moreover, the graph does not include the contributions of other terms in the sum, which add a relatively constant "background" that, in some cases, keeps $n > 1$ on both sides of the resonance.

If you agree to stay away from the resonances, the damping can be ignored, and the formula for the index of refraction simplifies:

$$n = 1 + \frac{Nq^2}{m\varepsilon_0} \sum_j \frac{f_j}{\omega_j^2 - \omega^2} \tag{F.147}$$

For most substances the natural frequencies ω_j are scattered all over the spectrum in a rather chaotic fashion. But for transparent materials, the nearest significant resonances typically lie in the ultraviolet, so that $gw < \omega_j$. In that case

$$\frac{1}{\omega_j^2 - \omega^2} = \frac{1}{\omega_j^2}\left(1 - \frac{\omega^2}{\omega_j^2}\right)^{-1} \cong \frac{1}{\omega_j^2}\left(1 + \frac{\omega^2}{\omega_j^2}\right)$$

and Eq. F.147 takes the form

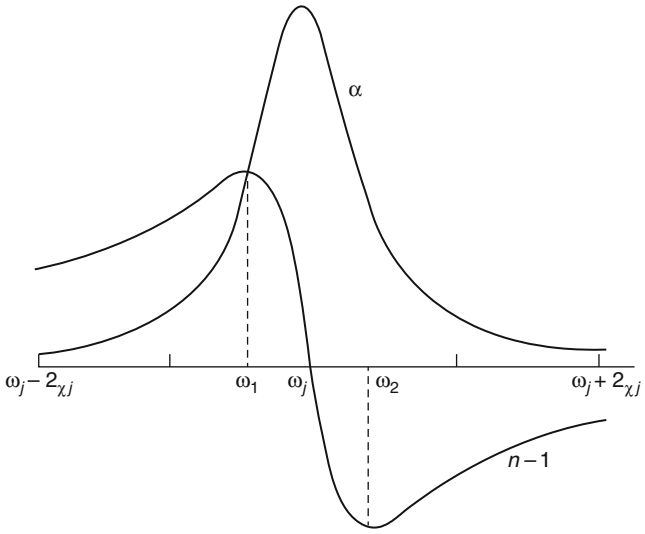


Fig. F.30 Plot of the index of refraction and absorption coefficient in the vicinity of one of the resonances [1]

$$n = 1 + \left(\frac{Nq^2}{m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^2}\right) + \omega^2\left(\frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^4}\right) \tag{F.148}$$

Or, in terms of the wavelength in vacuum ($\lambda = 2\pi c/\omega$):

$$n = 1 + A\left(1 + \frac{B}{\lambda^2}\right) \tag{F.149}$$

This is known as **Cauchy's formula**; the constant A is called the **coefficient of refraction** and B is called the **coefficient of dispersion**. Cauchy's equation applies reasonably well to most gases, in the optical region.

What Griffiths [1] has described in here is certainly not the complete story of dispersion in nonconducting media. Nevertheless, it does indicate how the damped harmonic motion of electrons can account for the frequency dependence of the index of refraction, and it explains why n is ordinarily a slowly increasing function of ω, with occasional "anomalous" regions where it precipitously drops.

Example 1: Assuming negligible damping ($\gamma_j = 0$), calculate the group velocity $\left(v_g = \frac{d\omega}{dk}\right)$ of the waves described by Eqs. F.141 and F.144. Show that $v_g < c$,even when $v > c$.

Solution: $k = \frac{\omega}{c}\left[1 + \frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\left(\omega_j^2 - \omega^2\right)^2}\right]$ and $v_g = \frac{d\omega}{dk} = \frac{1}{(dk/d\omega)}$

$$\frac{d\omega}{dk} = \frac{1}{c}\left[1 + \frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\left(\omega_j^2 - \omega^2\right)^2} + \omega\sum f_j \frac{-(-2\omega)}{\left(\omega_j^2 - \omega^2\right)^2}\right]$$

$$= \frac{1}{c}\left[1 + \frac{Nq^2}{2m\varepsilon_0}\sum f_j \frac{\left(\omega_j^2 + \omega^2\right)}{\left(\omega_j^2 + \omega^2\right)^2}\right]$$

$v_g = c\left[1 + \frac{Nq^2}{2m\varepsilon_0}\sum f_j \frac{\left(\omega_j^2 + \omega^2\right)}{\left(\omega_j^2 - \omega^2\right)^2}\right]$. Since the second term in square brackets is *positive*, it follows that $v_g < c$, whereas $v = \frac{\omega}{k} = c\left[1 + \frac{Nq^2}{2m\varepsilon_0}\sum \frac{f_j}{\left(\omega_j^2 - \omega^2\right)^2}\right]^{-1}$ is greater than c or less than c, depending on ω.

F.9 Electromagnetic Waves in Conductors

In our previous discussions of EM waves in matter, we have been assuming there are no free charges ($\rho_{free} = \rho_f$) or free currents ($\vec{J}_{free} = \vec{J}_f$); the only charges present were the bound charges in the dielectric. When dealing with conductors, there are plenty of electrons that are free to move about when an external electric field is applied. There are still bound charges in a conductor (atoms in a conductor generally only contribute one or two conduction electrons each), but we will find that (in a good conductor) the interaction of the fields with the free charges and currents will dominate over everything else.

In an ohmic materials which obeys Ohm's Law, we can write

$$\vec{J}(\vec{r}, t) = \sigma_c \vec{E}(\vec{r}, t) \tag{F.150}$$

where σ_c is the conductivity of the conducting materials (not to be mistaken for the surface charge density!) and $\sigma_c = 1/\rho_c$, which we will take to be uniform inside the conductor (ρ_c= resistivity of the metal conductor with dimension of Ohm-meter). For any free charge that might build up in the conductor due to these currents, we know from charge conservation:

$$\vec{\nabla} \cdot \vec{\boldsymbol{J}}(\vec{r},t) = -\frac{\partial \rho_f}{\partial t} \tag{F.151}$$

Thus inside such a conductor, we can assume that the linear/homogeneous/isotropic conducting medium has electric permittivity ε and magnetic permeability μ. Combine this with Gauss Law for electric fields in matter $\vec{\nabla} \cdot \left(\varepsilon \vec{\boldsymbol{E}}(\vec{r},t)\right) = -\rho_f$, to find how quickly the free charge distribution would dissipate. Bear in your mind that Maxwell's equations are always true, in matter or not; splitting things up into free and bound charges (or for that matter current) can be convenient for solving for the fields, but we do not have to do things this way. Now substitution of these relations into Eq. F.151, we can write:

$$\frac{\partial \rho_f}{\partial t} = \vec{\nabla} \cdot \left(\varepsilon \vec{\boldsymbol{E}}(\vec{r},t)\right) = -\frac{\sigma_c}{\varepsilon}\left[\left(\varepsilon \vec{\boldsymbol{E}}(\vec{r},t)\right)\right] = -\frac{\sigma_c}{\varepsilon}\rho_f \tag{F.152}$$

or

$$\frac{\partial \rho_f}{\partial t} = -\frac{\sigma_c}{\varepsilon}\rho_f \tag{F.153}$$

The solution to this first-order differential equation is:

$$\rho_f(t) = \rho(0)\exp(-\sigma_c t/\varepsilon) = \rho(\vec{r}, t=0)e^{-t/\tau_{relax}} \tag{F.154}$$

This is damped exponential type function. This also tells us that any free charge that might build up in a conductor will dissipate (due to the mutual repulsion of the charges) with a time constant $\tau_c = \varepsilon/\sigma_c$. The "time constant" represents the amount of time for the initial density to reduce to $1/e(\sim 37\%)$ of its original value. For a perfect conductor ($\sigma_c \to \infty$), the time constant goes to zero, meaning the charge density instantly dissipate. For a "good" conductor (like copper), where the conductivity is typically on the order of $10^8(\Omega.\mathrm{m})^{-1}$, the time constant is somewhere around 10^{-19} seconds, which is actually much smaller than the typical time between collisions of the electrons with the atoms making up the conductor, which is $\tau_c \tilde{} 10^{-14}$. For frequencies higher than $1/\tau_c$, Ohm's Law starts to break down, so τ_c is the important time constant here. Therefore, the assumption of dealing with the linear/homogeneous/isotropic conducting medium, is a valid assumption for the rest of this discussion and let us assume we are working with good conductors, and with frequencies that are at or below the optical range (10^{15}Hz, which is pushing the use of Ohm's law, but not so much that our results are no good). In other words, we should have $\omega \ll \sigma_c/\varepsilon$, or $\varepsilon\omega/\sigma_c \ll 1$. It turns out this assumption will make our calculation *much* easier as we go along.

Note that going for forward with further analysis of Electromagnetic Wave through Conduction we write the symbol of σ_c for conductivity of conducting

materials as σ_c, now that we know that is the case it is not indeed the surface charge density. Also we start using vector notation rather than bold character as well.

Now again assuming a conductor of linear, homogeneous and isotropic conducting material that has electric permittivity ε and magnetic permeability μ, we substitute $\rho_f = 0$, which means free charges dissipate very rather quick and $\vec{J}(\vec{r},t) = \sigma \vec{E}(\vec{r},t)$, which also means that $\vec{J}(\vec{r},t) \neq 0$, then we can write the Maxwell's Equations inside such conductor are as follow:

$$\vec{\nabla} \cdot \vec{E}(\vec{r},t) = \frac{\rho(\vec{r},t)}{\varepsilon}$$

$$\vec{\nabla} \cdot \vec{B}(\vec{r},t) = 0$$

$$\vec{\nabla} \times \vec{E}(\vec{r},t) = -\frac{\partial \vec{B}(\vec{r},t)}{\partial t}$$

and using Ohm's Law relation of $\vec{J}(\vec{r},t) = \sigma \vec{E}(\vec{r},t)$, we have the 4th Equation of Maxwell, as:

$$\vec{\nabla} \times \vec{B}(\vec{r},t) = \mu \vec{J}(\vec{r},t) + \mu\varepsilon \frac{\partial \vec{E}(\vec{r},t)}{\partial t} = \mu\sigma \vec{E}(\vec{r},t) + \mu\varepsilon \frac{\partial \vec{E}(\vec{r},t)}{\partial t}$$

(Solution of Eq. F.154)

Since electric charge is always conserved, thus the continuity equation inside the conductor is ends up with a homogeneous differential equation of first order as Eq. F.151b where the solution is provided as per Eq. F.154 and depicted in Fig. F.31, which is a damped exponential curve.

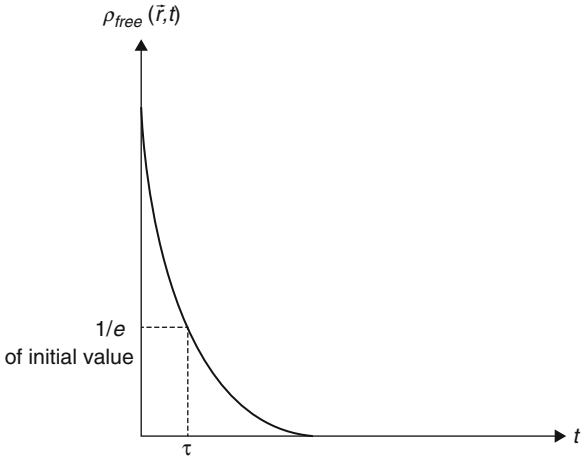


Fig. F.31 Sketch of Free Density Dissipated vs. Characteristic Time

Example 2: Calculate the charge relaxation time for copper

Solution: Assume the following data for cupper:

$$\rho_{Cu} = {}^{1}\!/_{\sigma_{Cu}} = 1.68 \times 10^{-8}\Omega - m \quad \Rightarrow \quad \sigma_{Cu} = {}^{1}\!/_{\rho_{Cu}} = 5.95 \times 10^{7} \quad \text{siemens/m}$$

If we assume $\varepsilon_{Cu} \approx 3\varepsilon_0 = 3 \times 8.85 \times 10^{-8} F/m$w for cupper metal, then:

$$\tau_{\text{Cu}}^{\text{relax}} = (\varepsilon_{Cu}/\sigma_{Cu}) = \rho_{Cu}\varepsilon_{Cu} = 4.5 \times 10^{-19}\,\text{sec}$$

However, knowing that the characteristic/mean collision time of free electros in pure copper is $\tau_{\text{Cu}}^{\text{coll}} \simeq \lambda_{\text{Cu}}^{\text{coll}}/v_{\text{thermal}}^{\text{Cu}}$ where $\lambda_{\text{Cu}}^{\text{coll}} \simeq 3.9 \times 10^{-8} m$ = mean free path (between successive collisions) in pure copper, and $v_{\text{thermal}}^{\text{Cu}} \simeq \sqrt{3k_B T/m_e} \simeq 12 \times 10^5 m/\,\text{sec}$ and thus we obtain $\tau_{\text{Cu}}^{\text{coll}} \simeq 3.2 \times 10^{-13}\,\text{sec}$.

Hence we see that the calculated charge relaxation time in pure copper, $\tau_{\text{Cu}}^{\text{relax}} \simeq 4.5 \times 10^{-19}\,\text{sec}$ is much less than the calculated collision in pure copper, $\tau_{\text{Cu}}^{\text{coll}} \simeq 3.2 \times 10^{-13}\,\text{sec}$. Furthermore, the experimentally measured charge relaxation time in pure copper is $\tau_{\text{Cu}}^{\text{relax}}(\text{experimental}) \simeq 4.0 \times 10^{-19}\,\text{sec}$, which is roughly 5 fold magnitude larger than the calculated charge relaxation time $\tau_{Cu}^{relax} \simeq 4.5 \times 10^{-19} sec$.

The problem here is that {the macroscopic} Ohm's Law is simply out of its range of validity on such short time scales! Two additional facts here are that both and C å ó are frequency-dependent quantities { i.e. $\varepsilon = \varepsilon(\omega)$ and $\sigma = \sigma(\omega)$}, which becomes increasingly important at the higher frequencies ($f = 2\pi/\omega \tilde{} 1/\tau_{relax}$) associated with short time-scale, transient-type phenomena!

So in reality, if we are willing to wait even a short time (e.g. $\Delta t \tilde{} 1ps = 10^{-12}\,\text{sec}$) then any initial free charge density $\rho_{free}(\vec{r}, t = 0)$ accumulated inside the conductor at $t = 0$ will have dissipated away/damped out, and from that time onwards, $\rho_{free}(\vec{r}, t) = 0$ *can* be safely assumed.

Thus, after many charge relaxation time constants, e.g. $20\tau_{relax} \leq \Delta t \simeq 1ps = 10^{-12}\,\text{sec}$, then

Maxwell's equations for a conductor become {with $\rho_{free}(\vec{r}, t \geq \Delta t) = 0$ from then onwards}:

New Sets of Maxwell's equation for a Charged-equilibrated Conductor

$$\vec{\nabla} \cdot \vec{E}(\vec{r}, t) = 0$$

$$\vec{\nabla} \cdot \vec{B}(\vec{r}, t) = 0$$

$$\vec{\nabla} \times \vec{E}(\vec{r}, t) = -\frac{\partial \vec{B}(\vec{r}, t)}{\partial t}$$

$$\vec{\nabla} \times \vec{B}(\vec{r},t) = \mu\vec{J}(\vec{r},t) + \mu\varepsilon\frac{\partial \vec{E}(\vec{r},t)}{\partial t} = \mu\left(\sigma\vec{E}(\vec{r},t) + \varepsilon\frac{\partial \vec{E}(\vec{r},t)}{\partial t}\right)$$

Now because these equations are different from the previous derivation(s) of monochromatic plane Electro-Magnetic (EM) waves propagating in free space/vacuum and/or in linear/homogeneous/isotropic nonconducting materials {only Eq. F.4) has changed}, we re-derive the wave equations for $\vec{E}$ and $\vec{B}$ from scratch. As before, we apply curl operation to Eqs. (F.3) and (F.4) and using the following vector identity, we get:

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{E}\right) = \left(\vec{\nabla} \cdot \vec{E}\right)\vec{\nabla} - \left(\vec{\nabla} \cdot \vec{\nabla}\right)\vec{E} + \left(\vec{E} \cdot \vec{\nabla}\right)\vec{\nabla} - \left(\vec{\nabla} \cdot \vec{\nabla}\right)\vec{E}$$

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{E}\right) = -\frac{\partial}{\partial t}\left(\vec{\nabla} \times \vec{B}\right)$$

$$= \vec{\nabla}\left(\vec{\nabla} \cdot \vec{E}\right)^{0} - \nabla^2\vec{E} = -\frac{\partial}{\partial t}\left(\mu\sigma\vec{E} + \mu\varepsilon\frac{\partial \vec{E}}{\partial t}\right)$$

$$= \vec{\nabla}2\vec{E} = \mu\varepsilon\frac{\partial^2 \vec{E}}{\partial t^2} + \mu\sigma\frac{\partial \vec{E}}{\partial t}$$

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{B}\right) = \mu\left[\sigma\left(\vec{\nabla} \times \vec{E}\right)\right] + \varepsilon\frac{\partial}{\partial t}\left(\vec{\nabla} \times \vec{E}\right)$$

$$= \vec{\nabla}\left(\vec{\nabla} \cdot \vec{B}\right) - \nabla^2\vec{B} = \mu\sigma\frac{\partial \vec{B}}{\partial t} - \mu\varepsilon\frac{\partial^2 \vec{B}}{\partial t^2}$$

$$= \vec{\nabla}^2\vec{B} = \mu\varepsilon\frac{\partial^2 \vec{B}}{\partial t^2} + \mu\sigma\frac{\partial \vec{B}}{\partial t}$$

Again, we can write:

$$\nabla^2\vec{E}(\vec{r},t) = \mu\varepsilon\frac{\partial^2 \vec{E}(\vec{r},t)}{\partial t^2} + \mu\sigma\frac{\partial \vec{E}(\vec{r},t)}{\partial t} \quad \text{and}$$

$$\nabla^2\vec{B}(\vec{r},t) = \mu\varepsilon\frac{\partial^2 \vec{B}(\vec{r},t)}{\partial t^2} + \mu\sigma\frac{\partial \vec{B}(\vec{r},t)}{\partial t}$$

Note that these 3D wave equations for $\vec{E}$ and $\vec{B}$ in a conductor have an additional term that has a single time derivative—which is analogous to a velocity-dependent damping term, e.g., for a mechanical harmonic oscillator.

The general solution(s) to the above wave equations are usually in the form of an oscillatory function * a damping term (i.e., a decaying exponential)—in the direction of the propagation of the EM wave, e.g., complex plane-wave type solutions for $\vec{E}$ and $\vec{B}$ associated with the above wave equation(s) are of the general form:

$$\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{i\left(\widetilde{k}x-\omega t\right)} \quad \text{and} \quad \widetilde{\vec{B}}(z,t)\widetilde{\vec{B}}_0 e^{\left(\widetilde{k}z-\omega t\right)} = \left(\frac{\widetilde{k}}{\omega}\right)\hat{k} \times \widetilde{\vec{E}}(z,t)$$

$$= \frac{1}{\omega}\hat{k} \times \widetilde{\vec{E}}(z,t) \tag{F.155}$$

With {frequency-dependent} complex wave number: $\widetilde{k}(\omega) = k(\omega) + i\kappa(w)$

Where $k(\omega) = Re\left(\widetilde{k}(\omega)\right)$ and $\kappa(w) = \Im m\left(\widetilde{k}(\omega)\right)$ and corresponding complex vector in the positive $+\hat{z}$ direction here is $\widetilde{\vec{k}}(\omega) = \widetilde{k}(\omega)\hat{k} = \widetilde{k}(\omega)\hat{z}$, i.e. $\widetilde{\vec{k}}(\omega) = [k(\omega) + i\kappa(\omega)]\hat{z}$.

We plug in $\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{i\left(\widetilde{k}x-\omega t\right)}$ and $\widetilde{\vec{B}}(z,t) = \widetilde{\vec{B}}_0 e^{i\left(\widetilde{k}z-\omega t\right)}$ into their respective wave equations above, and obtain from each wave equation the same/identical *characteristic equation* {dispersion relation} between complex $\widetilde{k}(\omega)$ and ω, we get the following relation as:

$$\widetilde{k}^2(\omega) = \mu\varepsilon\omega^2 + i\mu\sigma\omega \tag{F.156}$$

Thus, since $\widetilde{k}(\omega) = k(\omega) + i\kappa(\omega)$, then:

$$\boxed{\widetilde{k}^2(\omega) = [k(\omega) + i\kappa(\omega)]^2 = k^2(\omega) - \kappa^2(\omega) + 2ik(\omega)\kappa(\omega) = \mu\varepsilon\omega^2 + i\mu\sigma\omega} \tag{F.157}$$

If we temporarily suppress the ω-dependence of complex $\widetilde{k}(\omega)$, this relation becomes as:

$$\widetilde{k}^2(\omega) = (k + i\kappa)^2 = k^2 - \kappa^2 + 2ik\kappa = \mu\varepsilon\omega^2 + i\mu\sigma\omega \tag{F.158}$$

We can solve this relation to determine $(\omega) = Re\left(\widetilde{k}(\omega)\right)$ and $\kappa(\omega) = \mathfrak{T}m\left(\widetilde{k}(\omega)\right)$ as follows:

First, separate this relation into two relations by separating out its real and imaginary parts as:

$$\widetilde{k}^2(\omega) = (k + i\kappa) = k^2 - \kappa^2 + 2ik\kappa = i\mu\sigma\omega \tag{F.159}$$

Now we have two separate independent equations $k^2 - \kappa^2 = \mu\varepsilon\omega^2$ and $2k\kappa = \mu\sigma\omega$, and we have two unknowns such as k and κ. Hence, solving these two equations simultaneously we find the following results:

$$\begin{aligned} \kappa &= \frac{1}{2}\mu\sigma\omega/k \\ k^2 - k^2 &= k^2 - \left(\frac{1}{2}\mu\sigma\omega/k\right)^2 = k^2 = \frac{1}{k^2}\left(\frac{1}{2}\mu\sigma\omega\right)^2 = \mu\varepsilon\omega^2 \end{aligned} \tag{F.160}$$

Then multiply by k^2 and rearrange the terms to obtain the following relation:

$$k^4 - (\mu\varepsilon\omega^2)k^2 - \left(\frac{1}{2}\mu\sigma\omega\right)^2 = 0 \tag{F.161}$$

To solve this equation we let $x \equiv k^2$, $a \equiv 1$, $b \equiv -(\mu\omega\omega^2)$ and $c \equiv -\left(\frac{1}{2}\mu\sigma\omega\right)^2$, then Eq. F.161 reduces to the form of a quadratic equation of $ax^2 + bx + c = 0$ with solution roots of

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad \text{or} \quad k^2 = \frac{1}{2}\left[+(\mu\varepsilon\omega^2) \mp \sqrt{(\mu\varepsilon\omega^2)^2 + 4\left(\frac{1}{2}\mu\sigma\omega\right)^2}\right]$$

$$k = \frac{1}{2}(\mu\sigma\omega^2)\left[1 \mp \sqrt{1 + 4\frac{(\mu^2\sigma^2\omega^2)}{4(\mu^2\varepsilon^2\omega^4)}}\right] = \frac{1}{2}(\mu\sigma\omega^2)\left[1 \mp \sqrt{1 + \frac{(\sigma^2)}{(\varepsilon^2\omega^2)}}\right]$$

$$= \frac{1}{2}(\mu\sigma\omega^2) \times \left[1 \mp \sqrt{1 + \left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right]$$

Now we can see that on physical grounds ($k^2 > 0$), we ***must*** select the + sign, hence:

$$k^2 = \frac{1}{2}(\mu\sigma\omega^2)\left[1 + \sqrt{1 + \left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right]$$

and thus:

$$k = \sqrt{k^2} = \omega\sqrt{\frac{\varepsilon\mu}{2}\left[1 + \sqrt{1 + \left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right]^{1/2}} = \omega\sqrt{\frac{\varepsilon\mu}{2}\left[\sqrt{1 + \left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right]^{1/2}} \tag{F.162}$$

Having thus solved for k (or equivalently k^2), then we can use either of our original two relations to solve for κ, e.g. $k^2 - \kappa^2 = \mu\varepsilon\omega^2$, then:

$$\kappa^2 = k^2 - \mu\varepsilon\omega^2 = \frac{1}{2}(\mu\varepsilon\omega^2)\left[\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right] - \mu\varepsilon\omega^2\left[\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2} - 1\right]$$

Thus, we obtain

$$\begin{aligned} k(\omega) &= \Re e\left(\widetilde{k}(\omega)\right) = \omega\sqrt{\frac{\varepsilon\mu}{2}}\left[\sqrt{1+\left(\tfrac{\sigma}{\varepsilon\omega}\right)^2}+1\right]^{1/2} \\ \kappa(\omega) &= \Im m\left(\widetilde{k}(\omega)\right) = \omega\sqrt{\frac{\varepsilon\mu}{2}}\left[\sqrt{1+\left(\tfrac{\sigma}{\varepsilon\omega}\right)^2}-1\right]^{1/2} \end{aligned} \tag{F.163}$$

Note that the imaginary part of $\widetilde{k}$, $\kappa(\omega) = \Im m\left(\widetilde{k}(\omega)\right)$ results in an *exponential* attenuation and damping of the monochromatic plan EM wave with increasing z as:

$$\begin{aligned} &\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{-\kappa z} e^{i(kz-\omega t)} \quad \text{for Electric Field} \\ &\widetilde{\vec{B}}(z,t) = \widetilde{\vec{B}}_0 e^{-\kappa z} e^{i(kz-\omega t)} = \frac{1}{\omega}\widetilde{\vec{k}} \times \widetilde{\vec{E}}_0 e^{-\kappa z} e^{i(kz-\omega t)} \quad \text{for Magnetic Field} \end{aligned} \tag{F.164}$$

These solution (Eq. F.164) satisfy the above wave equations for *any* choice of $\widetilde{\vec{E}}_0$.

The characteristic distance over which $\vec{E}$ and $\vec{B}$ are attenuated, and reduced to $1/e = e^{-1} = 0.3679$ of their initial value at $z = 0$ is known as the *skin depth* and shown as $\delta_{skin\,depth} = \delta_{sc} \equiv 1/\kappa(\omega)$ and SI unit has dimension of meter.

$$\delta_{sc}(\omega) = \frac{1}{\kappa(\omega)} = \frac{1}{\omega\sqrt{\frac{\varepsilon\mu}{2}}\left[\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2}-1\right]^{1/2}} \Rightarrow \boxed{\begin{aligned} \widetilde{\vec{E}}(z=\delta_{sc},t) &= \widetilde{\vec{E}}_0\, e^{-1} e^{i(kz-\omega t)} \\ \widetilde{\vec{B}}(z=\delta_{sc},t) &= \widetilde{\vec{B}}_0 e^{-1} e^{i(kz-\omega t)} \end{aligned}}$$

The real part of $\widetilde{k}$, *i.e.* $k(\omega) = Re\left(\widetilde{k}(\omega)\right)$ determines the spatial wavelength $\lambda(\omega)$, the propagation speed v(ω) of the monochromatic ***EM*** plane wave (See Fig. F.26) in the conductor, and also the index of refraction:

$$\lambda(\omega) = \frac{2\pi}{k(\omega)} = \frac{2\pi}{Re\left(\widetilde{k}(\omega)\right)}$$

$$\mathrm{v}(\omega) = \frac{\omega}{k(\omega)} = \frac{\omega}{Re\left(\widetilde{k}(\omega)\right)}$$

$$n(\omega) = \frac{\omega}{\mathrm{v}(\omega)} = \frac{ck(\omega)}{\omega} = \frac{cRe\left(\widetilde{k}(\omega)\right)}{\omega}$$

The above plane wave solutions satisfy the above wave equations for any choice of $\widetilde{\vec{E}}_0$. As we have also seen before, it can similarly be shown here that Maxwell's Eqs. (F.1) and (F.2) $\{\nabla \cdot \vec{E} = 0$ and $\nabla \cdot \vec{B} = 0\}$ rule out the presence of any (longitudinal) z-components for $\vec{E}$ and $\vec{B}$ for EM waves propagating in the $+\hat{z}$-direction $\Rightarrow$ $\vec{E}$ and $\vec{B}$ are purely transverse wave.

If we consider, for example, a linearly polarized monochromatic plane EM wave propagating in the $+\hat{z}$-direction in a conducting medium, e.g., $\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{-\kappa z} e^{i(kz-\omega t)}\hat{x}$, then:

$$\begin{aligned}\widetilde{\vec{B}}(z,t) &= \frac{1}{\omega}\widetilde{\vec{k}} \times \widetilde{\vec{E}}(z,t) = \left(\frac{\widetilde{k}}{\omega}\right)\widetilde{E}_0 e^{-\kappa z} e^{i(kz-\omega t)}\hat{y} \\ &= \left(\frac{k+i\kappa}{\omega}\right)\widetilde{E}_0 e^{-\kappa z} e^{i(kz-\omega t)}\hat{y} \\ \Rightarrow \qquad & \widetilde{\vec{E}}(z,t) \perp \widetilde{\vec{B}}(z,t) \perp \not{2} \quad (+\not{2} = \text{propagation direction}) \end{aligned} \tag{F.165}$$

The complex wave-number $\widetilde{k} = k + ik = Ke^{i\phi_k}$ where: $K \equiv \left|\widetilde{k}\right| = \sqrt{k^2+\kappa^2}$ and $\phi_k \equiv \tan^{1}(\kappa/k)$, See Fig. F.32 below:

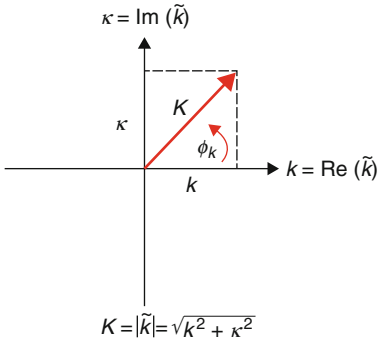


Fig. F.32 In the Complex $\widetilde{k}$ -plane

Then we see that:

$\widetilde{\vec{E}}(z,t) = \widetilde{\vec{E}}_0 e^{-\kappa z} e^{i(kz-\omega t)}\hat{x}$ has $\widetilde{E}_0 = E_0 e^{i\delta_E}$

and that: $\widetilde{\vec{B}}(z,t) = \widetilde{B}_0 e^{-\kappa z} e^{i(kz-\omega t)}\hat{y} = \frac{\widetilde{k}}{\omega}\widetilde{E}_0 e^{-\kappa z} e^{i(kz-\omega t)}\hat{y}$ has $\widetilde{B}_0 e^{-i\delta_B} = \frac{\widetilde{k}}{\omega}\widetilde{E}_0 = \frac{K e^{i\phi_k}}{\omega} E_0 e^{i\delta_E}$

Thus, we see that: $B_0 e^{-i\delta_B} = \frac{K e^{i\phi_k}}{\omega} E_0 e^{i\delta_E} = \frac{K}{\omega} E_0 e^{i(\delta_E+\phi_k)} = \frac{\sqrt{k^2+\kappa^2}}{\omega} E_0 e^{i(\delta_E+\phi_k)}$

i.e., inside a conductor, $\vec{E}$ and $\vec{B}$ are no longer in phase with each other.

Phases of $\vec{E}$ and $\vec{B}$: $\boxed{\delta_B = \delta_E + \phi_k}$

We also see that: $\delta\varphi_{B-E} \equiv \delta_B - \delta_E = \phi_k$ magnetic field ***lags*** behind electric field.

We also see that: $\boxed{\frac{B_0}{E_0} = \frac{K}{\omega} = \left[\varepsilon\mu\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right]^{1/2} \neq \frac{1}{c}}$

The real physical $\vec{E}$ and $\vec{B}$ fields associated with linearly polarized monochromatic plane *EM* waves propagating in a conducting medium are *exponentially* damped:

$$\vec{E}(z,t) = Re\left(\widetilde{\vec{E}}(z,t)\right) = E_0 e^{-\kappa z}\cos(kz - \omega t + \delta_E)\hat{x}$$

$$\vec{B}(z,t) = Re\left(\vec{B}(z,t)\right) = B_0 e^{-\kappa z}\cos(kz - \omega t + \delta_B)\hat{y}$$

$$= B_0\cos(kz - \omega t + \{\delta_E + \phi_k\})\hat{y}$$

$$\frac{B_0}{E_0} = \frac{K(\omega)}{\omega} = \left[\varepsilon\mu\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right]^{1/2} \tag{F.166}$$

$$K(\omega) \equiv \left|\hat{k}(\omega)\right| = \sqrt{k^2(\omega)+\kappa^2(\omega)} = \omega\left[\varepsilon\mu\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2}\right]^{1/2}$$

$$\delta_B = \delta_E + \phi_k \qquad \widetilde{\vec{k}}(\omega) = [k(\omega) + i\kappa(\omega)]\hat{z}$$

$$\phi_k(\omega) \equiv \tan^{-1}\left(\frac{\kappa(\omega)}{k(\omega)}\right) \quad \text{and} \quad \widetilde{k}(\omega) = \left|\widetilde{\vec{k}}(\omega)\right| = k(\omega) + i\kappa(\omega)$$

Definition of the skin depth in a conductor:

$$\delta_{sc}(\omega) \equiv \frac{1}{\kappa(\omega)} = \frac{1}{\omega\sqrt{\frac{\varepsilon\mu}{2}\left[\sqrt{1+\left(\frac{\sigma}{\varepsilon\omega}\right)^2}-1\right]^{1/2}}} =$$

Distance over which the $\vec{E}$ and $\vec{B}$ fields fall to $1/e = e^{-1} = 0.3679$ of their initial values

Example 3: What is the skin depth of fine silver at microwave frequency of 1010 HZ, which is common microwave region; assume the silver has conductivity of $g = 10^{10}\text{HZ}$

Solution: The skin depth is presented by the following equation as:

$$\delta = \sqrt{\frac{2}{\mu_0 \omega g}} = \sqrt{\frac{2}{(2\pi \times 10^{10})(4\pi \times 10^{-7})(3 \times 10^3)}} = 9.2 \times 10^{-5}\,\text{cm}$$

Thus at microwave frequencies the skin depth in silver is very small, and consequently, the difference in performance between a pure silver component and a silver-plated brass component would be expected to be negligible.

Example 4: For seawater case, we calculate the frequency at which the skin depth is one meter. For seawater, $\mu = \mu_0$ and $g \approx 4.3\,\text{S/m}$.

Solution: The expression for the frequency corresponding to a given skin depth δ is:

$$\omega = \frac{2}{g\mu_0\delta^2} = \frac{2}{4.3 \times 4\pi \times 10^{-7}\delta^2} = \frac{3.70 \times 10^5}{\delta^2}(\sec)^{-1}$$

which yields:

$$f = 58 \times 10^3\,\text{HZ}$$

or frequency of 60 kHz for a skin depth of one meter. If a submarine is equipped with a very sensitive receiver and if a very powerful transmitter is used, it is possible to communicate with a submerged submarine. However, a low radiofrequency must be used, and even then an extremely sever attenuation of the signal occurs. At five skin depths (5 meter in the case calculated above), only 1 percent of the initial electric field remains and only 0.01 percent of the incident power.

F.10 References

1. David Griffiths “Introduction to electrodynamics”, 3rd Edition, Prentice Hall, 1999.
2. Leonard Eyges, “The Classical Electromagnetic Field” Dover Publication, 1972.
3. J. R. Reitz, F. J. Milford, and R. W. Christy, “Foundations of Electromagnetic Theory”, 3rd Edition, Section 17-5, Addison Wesley, 1979.
4. M. B. James and D. J. Griffiths, *Am. J. Physics*, **60**, 309 (1992).
5. N. Ashby, *Am. J. Phys.* **43**, 553 (1975).
6. H. C. Ohanian, *Am. J. Phys.* **51**, 1020 (1983).
7. Handbook of Chemistry and Physics, 78th edition. (Boca Raton: CRC Press, Inc., 1997).
8. Francis A. Jenkins, “Fundamentals of Optics”, McGraw-Hill Science/Engineering/Math; 4th edition December 3, 2001.

Appendix G
Short Course in Optics

One of the major goals of physics and so as ours in case of laser in particular and its interaction with materials is to understand the nature of light. Due to the complexity nature of the light this goal is very difficult to fully achieve but this complication means that light offers many opportunities for different applications including optical interferences and in our case response of materials to laser radiation and its interaction with matter and specifically metallic materials.

G.1 Light as a Wave

It is a great advantages that if we can account for the laws of reflection and refraction in terms of waves and give some physical meaning to the index of refraction and then later on be able to tight them down to Maxwell's electromagnetic theory for the purpose of laser interaction with materials where we need to understand response of materials to laser radiation. Huygens' wave theory is based on a geometrical construction that allows us to tell where a given wave front will be at any time in the future if we know its present position. This construction is based on **Huygens Principle**, which is:

> All points on a wave front serve as point sources of spherical secondary wavelets. After a time t, the new position of the wave front will be that of a surface to these secondary wavelets.

A simple example of Huygens' Principle can be presented by Fig. G.1 where the present location of a wavefront of a plane wave traveling to the right in vacuum is represented by plane *ab*, perpendicular to the page[22]. Then question is, where will the wavefront be at time Δt later?.

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

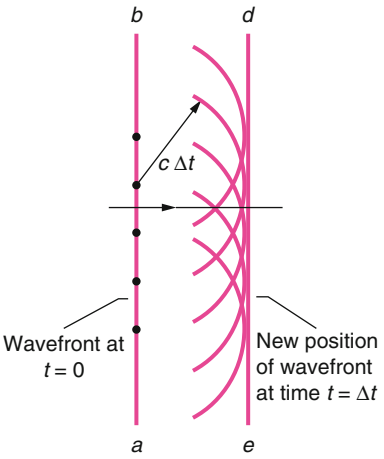


Fig. G.1 The propagation of a plane wave in vacuum, as portrayed by Huygens' principle [1]

We let several points on plane *ab* (the dots) serve as sources of spherical secondary wavelets that are emitted at $t = 0$. At time Δt, the radius of all these spherical wavelets will grown to $c\Delta t$, where c is the speed of light in vacuum. We draw plane *de* tangent to these wavelets at time Δt. This plane represents the wavefront of the plane wave at time Δt; it is parallel to plane *ab* and a perpendicular distance $c\Delta t$ from it.

G.2 Refraction of Light

Another phenomenon which can be analyzed using the Huygens principle is the refraction of light—that is, the change in direction of a light beam it passes from one medium to another in which its speed is different. This can be evaluated with the help of the wave front diagram of Fig. G.2. At $t = 0$ the wave front $\underline{he}$ just comes in contact with the boundary between the two media. We suppose that the speed of light, v_2, in the new medium is less than that in the first one, so that secondary wavelets generated at the interface travel a shorter distance in the same time interval than do wavelets in the first medium. At the moment $t = 0$ when the end $\underline{h}$ of the wave front $\underline{he}$ reaches the boundary, end $\underline{e}$ is still the distance $\underline{ec}$ away.

Since the speed of light in the first medium is v_1, end $\underline{e}$ requires the time $t = \frac{ec}{v_1}$ to reach the boundary at $\underline{c}$. In this period of time the end $\underline{h}$ of the wave front proceeds to $\underline{g}$, where $\underline{hg}$ is smaller than $\underline{ec}$. Evidently

$$ec = v_1 t \quad \text{and} \quad hg = v_2 t$$

The angle θ_1 between an approaching wave front and the boundary between two media is called the *angle of incident* of the wave front; this angle is equal to that made by an approaching ray with the normal to the boundary as in Fig. G.3. The

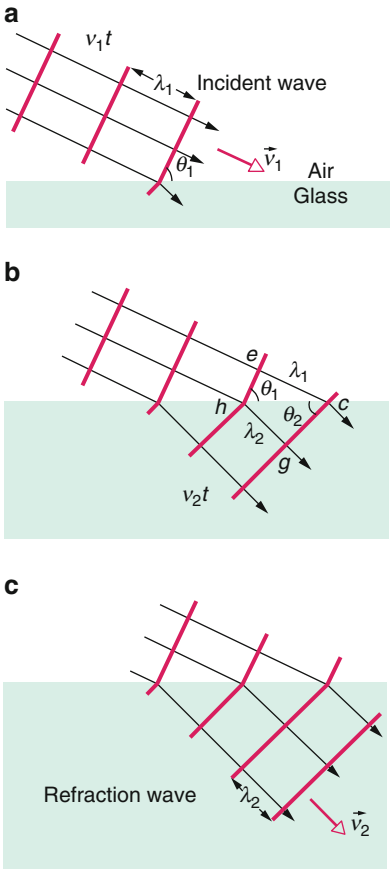


Fig. G.2 The refraction of a plane wave at media 1 and 2 interface, as portrayed by Huygens' principal. Parts (**a**) through (**c**) represent three successive stages of the refraction [1]

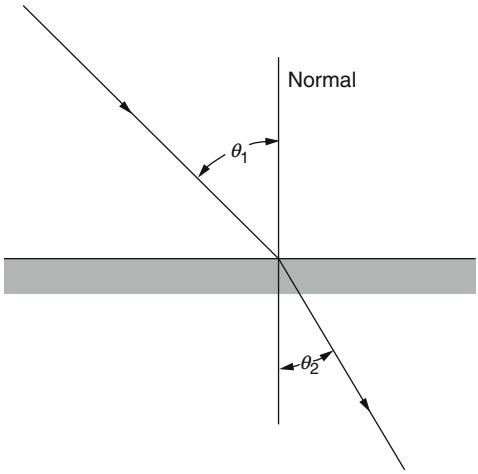


Fig. G.3 Ray treatment of refraction [2]

angle θ_2 between a receding wave front and the boundary between two media through which it has passed is called the *angle of refraction*; this angle is equal to that made by a receding ray with the normal.

From Fig. G.2 we see that

$$\sin\theta_1 = \frac{ec}{hc} = \frac{v_1 t}{hc}$$

and

$$\sin\theta_2 = \frac{hg}{hc} = \frac{v_2 t}{hc}$$

So that

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} \tag{G.1}$$

From Eq. (G.1) we can conclude that the *ratio of the sines of the angles of incidence and refraction is equal to the ratio of the speeds of light in the two media*. Equation (G.1) is known as *Snell's law* after its discoverer, the seventeenth-century Dutch astronomer Willebrord Snell.

The ratio between the speed of light c in free space and its speed v in a particular medium is called the *index of refraction of the medium*, the symbol for which is n and that is written as follows:

$$n = \frac{c}{v} \tag{G.2}$$

Since $v_1 = c/n_1$ and $v_2 = c/n_2$, where n_1 and n_2 are the indexes of refraction of the two media, Snell's law, Eq. (G.1) can be written in the alternative form as;

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{G.3}$$

Table G.1 is a list of the values of n for a number of substances. The greater the index of refraction, the greater the extent to which a light is deflected upon entering or leaving that medium.

G.3 Wave Equation

The wave equation is an important second-order linear partial differential equation of waves, such as sound waves, light waves and water waves. It arises in fields such as acoustics, electromagnetics, and fluid dynamics. Historically, the problem of a vibrating string such as that of a musical instrument was studied by Jean le Rond

Table G.1 Indexes of refraction for number of substances

Substance	n	Substance	n
Air	1.0003	Glass, flint	1.63
Benzene	1.50	Glycerin	1.47
Carbon disulfide	1.63	Ice	1.31
Diamond	2.42	Quartz	1.46
Ethyl alcohol	1.36	Water	1.34
Glass crown	1.52		

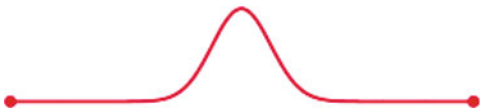


Fig. G.4 A pulse traveling through a string with fixed endpoints as modeled by the wave equation [3]

d'Alembert, Leonhard Euler, Daniel Bernoulli, and Joseph-Louis Lagrange (Fig. G.4).

The wave equation is the prototypical example of a hyperbolic partial differential equation. In its simplest form, the wave equation refers to a scalar function $u = (x_1, x_2, \ldots, x_n, t)$ that satisfies:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u \tag{G.4}$$

where ∇^2 is the (spatial) Laplacian and where c is a fixed constant equal to the propagation speed of the wave. This is known as the non-dispersive wave equation. For a sound wave in air at 20°C this constant is about 343 m/s (see speed of sound). For the vibration of a string the speed can vary widely, depending upon the linear density of the string and the tension on it. For a spiral spring (a slinky) it can be as slow as a meter per second. More realistic differential equations for waves allow for the speed of wave propagation to vary with the frequency of the wave, a phenomenon known as dispersion. In such a case, c must be replaced by the *phase velocity*:

$$v_p = \frac{\omega}{k} \tag{G.5}$$

Where

ω = Angular Frequency.
k = Wave number

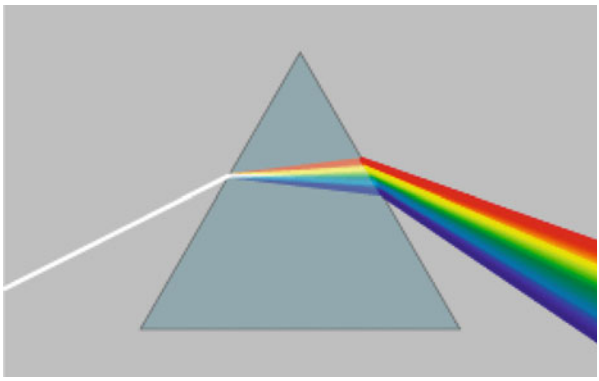


Fig. G.5 In a prism, material dispersion (a wavelength-dependent refractive index) causes different colors to refract at different angles, splitting white light into a rainbow [3]

G.3.1 Dispersion in Optics

In optics, dispersion is the phenomenon in which the phase velocity of a wave depends on its frequency, or alternatively when the group velocity depends on the frequency. Media having such a property are termed dispersive media. Dispersion is sometimes called chromatic dispersion to emphasize its wavelength-dependent nature, or group-velocity dispersion (GVD) to emphasize the role of the group velocity.

The most familiar example of dispersion is probably a rainbow, in which dispersion causes the spatial separation of a white light into components of different wavelengths (different colors). However, dispersion also has an effect in many other circumstances: for example, GVD causes pulses to spread in optical fibers, degrading signals over long distances; also, a cancellation between group-velocity dispersion and nonlinear effects leads to soliton waves. Dispersion is most often described for light waves, but it may occur for any kind of wave that interacts with a medium or passes through an inhomogeneous geometry (e.g., a waveguide), such as sound waves (Fig. G.5).

There are generally two sources of dispersion: material dispersion and waveguide dispersion. Material dispersion comes from a frequency-dependent response of a material to waves. For example, material dispersion leads to undesired chromatic aberration in a lens or the separation of colors in a prism. Waveguide dispersion occurs when the speed of a wave in a waveguide (such as an optical fiber) depends on its frequency for geometric reasons, independent of any frequency dependence of the materials from which it is constructed. More generally, "waveguide" dispersion can occur for waves propagating through any inhomogeneous structure (e.g., a photonic crystal), whether or not the waves are confined to some region. In general, both types of dispersion may be present, although they are not strictly additive. Their combination leads to signal degradation in optical fibers for telecommunications, because the varying delay in arrival time between different components of a signal "smears out" the signal in time.

G.3.2 Material Dispersion in Optics

Material dispersion can be a desirable or undesirable effect in optical applications. The dispersion of light by glass prisms is used to construct spectrometers and spectroradiometers. Holographic gratings are also used, as they allow more accurate discrimination of wavelengths. However, in lenses, dispersion causes chromatic aberration, an undesired effect that may degrade images in microscopes, telescopes and photographic objectives.

The phase velocity, v, of a wave in a given uniform medium is given by

$$v = \frac{c}{n}$$

where c is the speed of light in a vacuum and n is the refractive index of the medium.

In general, the refractive index is some function of the frequency f of the light, thus $n = n(f)$, or alternatively, with respect to the wave's wavelength $n = n(?)$. The wavelength dependence of a material's refractive index is usually quantified by an empirical formula, the Cauchy or Sellmeier equations.

Because of the Kramers–Kronig relations, the wavelength dependence of the real part of the refractive index is related to the material absorption, described by the imaginary part of the refractive index (also called the *extinction coefficient*). In particular, for non-magnetic materials ($\mu = \mu_0$), the susceptibility χ that appears in the Kramers–Kronig relations is the electric susceptibility $\chi_e = n^2 - 1$.

The most commonly seen consequence of dispersion in optics is the separation of white light into a color spectrum by a prism. From Snell's law it can be seen that the angle of refraction of light in a prism depends on the refractive index of the prism material. Since that refractive index varies with wavelength, it follows that the angle that the light is refracted by will also vary with wavelength, causing an angular separation of the colors known as *angular dispersion*.

For visible light, most transparent materials (e.g., glasses) have:

$$1 < n(\lambda_{\text{red}}) < n\left(\lambda_{\text{yellow}}\right) < n(\lambda_{\text{blue}})$$

or alternatively:

$$\frac{\mathrm{d}n}{\mathrm{d}\lambda} < 0$$

that is, refractive index n decreases with increasing wavelength λ. In this case, the medium is said to have normal dispersion. Whereas, if the index increases with increasing wavelength the medium has anomalous dispersion (Fig. G.6).

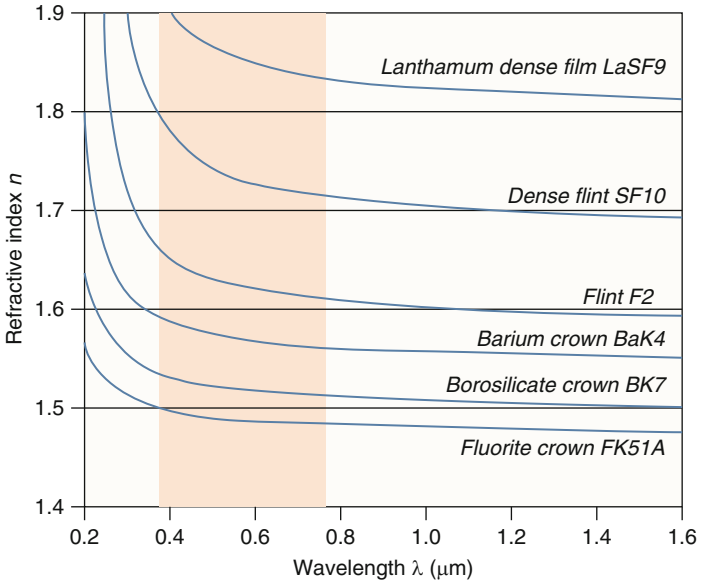


Fig. G.6 The variation of refractive index vs. wavelength for various glasses. The wavelengths of visible light are shaded in *red*

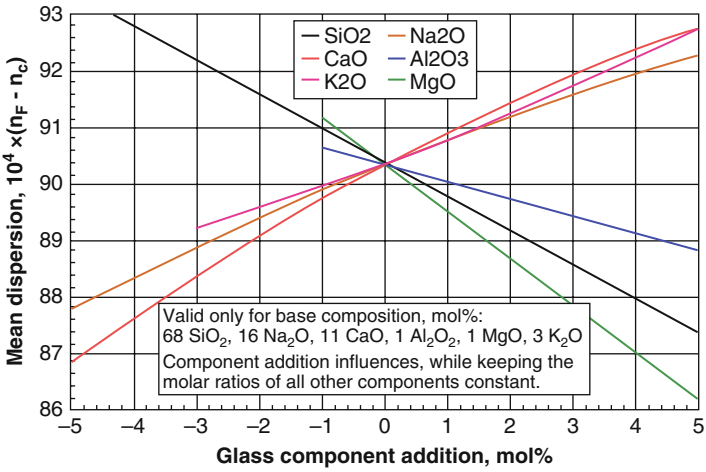


Fig. G.7 Influences of selected glass component additions on the mean dispersion of a specific base glass (n_F valid for $\lambda = 486$ nm (*blue*), n_C valid for $\lambda = 656$ nm (*red*))

At the interface of such a material with air or vacuum (index of ~1), Snell's law predicts that light incident at an angle θ to the normal will be refracted at an angle $\arcsin\big(\sin(\theta)/n$. Thus, blue light, with a higher refractive index, will be bent more strongly than red light, resulting in the well-known rainbow pattern (Fig. G.7).

G.3.3 Phase Velocity

The phase speed (or phase velocity when considered as a vector) of a wave is the rate at which the phase of the wave propagates in space. This is the speed at which the phase of any one frequency component of the wave travels. For such a component, any given phase of the wave (for example, the crest) will appear to travel at the phase speed. The phase speed is given in terms of the wavelength λ (lambda) and period T as

$$v_p = \frac{\lambda}{T}$$

Or, equivalently, in terms of the wave's angular frequency ω and wave number k by

$$v_p = \frac{\omega}{k}$$

In a dispersive medium, the phase speed varies with frequency and is not necessarily the same as the group speed of the wave, which is the rate that changes in amplitude (known as the envelope of the wave) propagate.

The phase speed of electromagnetic radiation may, under certain circumstances, (for example anomalous dispersion) exceed the speed of light in a vacuum, but this does not indicate any superluminal information or energy transfer. It was theoretically described by physicists such as Arnold Summerfield and Léon Brillouin (Fig. G.8).

G.3.4 Group Velocity

The group velocity of a wave is the velocity with which the overall shape of the wave's amplitudes—known as the modulation or envelope of the wave—propagates through space.

For example, imagine what happens if a stone is thrown into the middle of a very still pond. When the stone hits the surface of the water, a circular pattern of waves appears. It soon turns into a circular ring of waves with a quiescent center. The ever expanding ring of waves is the wave group, within which one can discern individual wavelets of differing wavelengths traveling at different speeds. The longer waves travel faster than the group as a whole, but they die out as they approach the leading

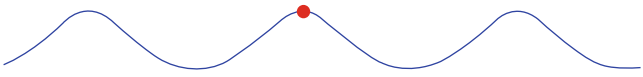


Fig. G.8 Phase velocity in periodic gravity waves on the surface of deep water. The *red dot* moves with the phase velocity, and is located at a fixed wave phase: the crest for the case shown [3]

edge. The shorter waves travel slower and they die out as they emerge from the trailing boundary of the group.

The group velocity v_g is defined by the equation

$$v_g = \frac{\partial \omega}{\partial k}$$

where

$\omega =$ is the wave's angular frequency;

$k =$ is the wave number.

The function $\omega(k)$, which gives ω as a function of k, is known as the dispersion relation. If ω is directly proportional to k, then the group velocity is exactly equal to the phase velocity. Otherwise, the envelope of the wave will become distorted as it propagates. This "group velocity dispersion" is an important effect in the propagation of signals through optical fibers and in the design of high-power, short-pulse lasers.

Note: The above definition of group velocity is only useful for wave packets, which is a pulse that is localized in both real space and frequency space. Because waves at different frequencies propagate at differing phase velocities in dispersive media, for a large frequency range (a narrow envelope in space) the observed pulse would change shape while traveling, making group velocity an unclear or useless quantity (Fig. G.9).

Note: The red dot moves with the phase velocity, and the green dots propagate with the group velocity. In this deep-water case, the phase velocity is twice the group velocity. The red dot overtakes two green dots, when moving from the left to the right of the figure.

New waves seem to emerge at the back of a wave group, grow in amplitude until they are at the center of the group, and vanish at the wave group front. For surface gravity waves, the water particle velocities are much smaller than the phase velocity, in most cases.

G.3.4.1 Physical Interpretation

The group velocity is often thought of as the velocity at which energy or information is conveyed along a wave. In most cases this is accurate, and the group velocity can be thought of as the signal velocity of the waveform. However, if the wave is travelling through an absorptive medium, this does not always hold. Since the 1980s, various experiments have verified that it is possible for the group velocity of laser light pulses sent through specially prepared materials to significantly

Fig. G.9 Frequency dispersion in groups of gravity waves on the surface of deep water [3]

exceed the speed of light in vacuum. However, superluminal communication is not possible in this case, since the signal velocity remains less than the speed of light. It is also possible to reduce the group velocity to zero, stopping the pulse, or have negative group velocity, making the pulse appear to propagate backwards. However, in all these cases, photons continue to propagate at the expected speed of light in the medium.

Anomalous dispersion happens in areas of rapid spectral variation with respect to the refractive index. Therefore, negative values of the group velocity will occur in these areas. Anomalous dispersion plays a fundamental role in achieving backward propagating and superluminal light. Anomalous dispersion can also be used to produce group and phase velocities that are in different directions. Materials that exhibit large anomalous dispersion allow the group velocity of the light to exceed c and/or become negative[23].

G.3.4.2 History

The idea of a group velocity distinct from a wave's phase velocity was first proposed by W.R. Hamilton in 1839, and the first full treatment was by Rayleigh in his "Theory of Sound" in 1877[24].

G.3.4.3 Matter-Wave Group Velocity

Albert Einstein first explained the wave–particle duality of light in 1905. Louis de Broglie hypothesized that any particle should also exhibit such a duality. The velocity of a particle, he concluded then (but may be questioned today, see above), should always equal the group velocity of the corresponding wave. De Broglie deduced that if the duality equations already known for light were the same for any particle, then his hypothesis would hold. This means that

$$v_g = \frac{\partial \omega}{\partial k} = \frac{\partial (E/\hbar)}{\partial (p/\hbar)} = \frac{\partial E}{\partial p}$$

where

E is the total energy of the particle,
p is its momentum,
$\hbar$ is the reduced Planck constant.
For a free non-relativistic particle it follows that

$$v_g = \frac{\partial E}{\partial p} = \frac{\partial}{\partial p}\left(\frac{1}{2}\frac{p^2}{m}\right) = \frac{p}{m} = v$$

where

m is the mass of the particle and

v its velocity.

Also in special relativity we find that

$$\begin{aligned} v_g &= \frac{\partial E}{\partial p} = \frac{\partial}{\partial p}\left(\sqrt{p^2c^2 + m^2c^4}\right) \\ &= \frac{pc^2}{\sqrt{p^2c^2 + m^2c^4}} \\ &= \frac{p}{m\sqrt{(p/(mc))^2 + 1}} \\ &= \frac{p}{m\gamma} \\ &= \frac{mv\lambda}{m\gamma} \\ &= v \end{aligned}$$

where

m is the rest mass of the particle,

c is the speed of light in a vacuum,

γ is the Lorentz factor.

and v is the velocity of the particle regardless of wave behavior.

Group velocity (equal to an electron's speed) should not be confused with phase velocity (equal to the product of the electron's frequency multiplied by its wavelength).

Both in relativistic and non-relativistic quantum physics, we can identify the group velocity of a particle's wave function with the particle velocity. Quantum mechanics has very accurately demonstrated this hypothesis, and the relation has been shown explicitly for particles as large as molecules.

G.4 Angular Frequency

In physics, angular frequency ω (also referred to by the terms angular speed, radial frequency, circular frequency, orbital frequency, and radian frequency) is a scalar measure of rotation rate. Angular frequency (or angular speed) is the magnitude of the vector quantity angular velocity. The term angular frequency vector $\vec{\omega}$ is sometimes used as a synonym for the vector quantity angular velocity

In SI units, angular frequency is measured in radians per second, with units s^{-1} since radians are unit less.

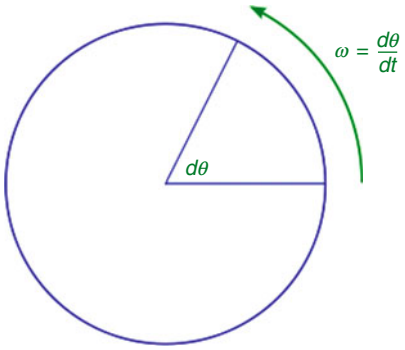


Fig. G.10 Angular frequency is a measure of how fast an object is rotating around its axis

One revolution is equal to 2π radians, hence

$$\omega = \frac{2\pi}{T} = 2\pi f = \frac{|v|}{|r|} \tag{G.6}$$

Where

$\omega =$ is the angular frequency or angular speed (measured in radians per second),

$T =$ is the period (measured in seconds),

$f =$ is the ordinary frequency (measured in hertz),

$v =$ is the tangential velocity of a point about the axis of rotation (measured in meters per second),

$r =$ is the radius of rotation (measured in meters).

Angular frequency is therefore a simple multiple of ordinary frequency. However, using angular frequency is often preferable in many applications, as it avoids the excessive appearance of π. In fact, it is used in many fields of physics involving periodic phenomena, such as quantum mechanics and electrodynamics (Fig. G.10).

For example:

$$a = -\omega^2 x$$

Using 'ordinary' revolutions-per-second frequency, this equation would be:

$$a = -4\pi^2 f^2 x$$

Another often encountered expression when dealing with small oscillations or where damping is negligible is:

$$\omega^2 = \frac{k}{m}$$

Where

$k =$ is the spring constant.

$m =$ is the mass of the object.

This is referred to as the natural frequency.

Angular frequency inside an LC circuit can also be defined as the square root of the inverse of capacitance (measured in farads), times the inductance of the circuit (in henrys)

$$\omega = \sqrt{\frac{1}{LC}}$$

G.5 Wave Number

Wave number is in the physical sciences a property of a wave proportional to the reciprocal of the wavelength. It can be defined as either

the number of wavelengths per unit distance, that is, $1/\lambda$ where $\lambda =$ wavelength,

or alternatively as $2\pi/\lambda$, sometimes termed the angular wave number or circular wave number or, simply wave number.

For electromagnetic radiation, wave number is proportional to frequency and to photon energy.

Because of this, wave numbers are used as a unit of energy in spectroscopy. In the SI units, wave number is expressed in units of reciprocal meters (m^{-1}), but in spectroscopy it is usual to give wave numbers in inverse centimeters (cm^{-1}). For the special case of an electromagnetic wave,

$$k \equiv \frac{2\pi}{\lambda} = \frac{2\pi\nu}{v_p} = \frac{\omega}{v_p} = \omega\sqrt{\mu\varepsilon} = \frac{E}{\hbar c}$$

Where ν is the frequency of the wave, v_p is the phase velocity of the wave (if it travels in vacuum, $v_p = c$), ω is the angular frequency of the wave, E is the energy of the wave, $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant, and c is the velocity of light in vacuum. The angular wave number is the magnitude of the wave vector.

For the special case of a matter wave, for example an electron wave, in the non-relativistic approximation:

$$k \equiv \frac{2\pi}{\lambda} = \frac{p}{\hbar} = \frac{\sqrt{2mE}}{\hbar}$$

Here p is the momentum of the particle, m is the mass of the particle, E is the kinetic energy of the particle, and $\hbar$ is the reduced Planck's constant.

G.6 References

1. Holliday D, Resnick R, Walker J. (2005) Fundamentals of physics. 7th edition, John Wiley

2. Jenkins FA (2001) Fundamentals of optics. McGraw-Hill Science/Engineering/Math; 4th edition
3. Wave Equation—Wikipedia, the free encyclopedia

Appendix H
Short Course in Heat Conduction

One of the major goals of physics and so as ours in case of laser in particular and its interaction with materials is to understand the nature of light. Due to the complexity nature of the light this goal is very difficult to fully achieve but this complication means that light offers many opportunities for different applications including optical interferences and in our case response of materials to laser radiation and its interaction with matter and specifically metallic materials. This leads to dealing with heat conduction equation in a very complex form, therefore we need to deal with Heat Transport from aspect of *Conduction*, *Convection* and *Radiation*.

H.1 Fourier Law of Heat Conduction

When there exists a temperature gradient within a body, heat energy will flow from the region of high temperature to the region of low temperature. This phenomenon is known as conduction heat transfer, and is described by **Fourier's Law** (named after the French physicist Joseph Fourier),

$$\mathrm{q} = -k\vec{\nabla} T$$

This equation determines the heat flux vector **q** for a given temperature profile T and thermal conductivity k. The minus sign ensures that heat flows down the temperature gradient.

Heat Equation (Temperature Determination)

The temperature profile within a body depends upon the rate of its internally-generated heat, its capacity to store some of this heat, and its rate of thermal conduction to its boundaries (where the heat is transferred to the surrounding environment). Mathematically this is stated by the **Heat Equation**,

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

$$\vec{\nabla}^2 T - \frac{1}{\alpha}\frac{\partial T}{\partial t} = -\frac{1}{k} q_{\text{gen}}$$

along with its boundary conditions, equations that prescribe either the temperature T on, or the heat flux q through, all of the body boundaries

$$T(\Omega_a) = T_{\text{prescribed}}$$

$$q(\Omega_b) = q_{\text{prescribed}}$$

$$\Omega_a \cup \Omega_b = \Omega$$

In the Heat Equation, the power generated per unit volume is expressed by q_{gen}. The thermal diffusivity α is related to the thermal conductivity k, the specific heat c, and the density ρ by,

$$\alpha = \frac{k}{\rho c}$$

For **Steady State** problems, the Heat Equation simplifies to,

$$\vec{\nabla}^2 T = -\frac{1}{k} q_{\text{gen}}$$

Derivation of the Heat Equation

The heat equation follows from the **conservation of energy** for a small element within the body,

heat conducted in	+	heat generated within	=	heat conducted	+	change in energy stored within

We can combine the heats conducted in and out into one "net heat conducted out" term to give,

Net heat conducted out	=	heat generated within	–	change in energy stored within

Mathematically, this equation is expressed as,

$$\vec{\nabla}^2 \cdot \mathbf{q} = q_{\text{gen}} - \frac{de}{dt}$$

The change in internal energy e is related to the body's ability to store heat by raising its temperature, given by,

$$\frac{de}{dt} = \rho c \frac{dT}{dt}$$

One can substitute for **q** using Fourier's Law of heat conduction from above to arrive at the Heat Equation,

$$\vec{\nabla} \cdot \left(-k \vec{\nabla} T \right) = q_{\text{gen}} - \rho c \frac{dT}{dt}$$

$$-k \vec{\nabla}^2 T + \rho c \frac{\partial T}{\partial t} = q_{\text{gen}}$$

$$\vec{\nabla}^2 T - \frac{1}{\alpha} \frac{\partial T}{\partial t} = -\frac{1}{k} q_{\text{gen}}$$

Blackbody	A body with a surface emissivity of 1. Such a body will emit all of the thermal radiation it can (as described by theory), and will absorb 100 % of the thermal radiation striking it. Most physical objects have surface emissivities less than 1 and hence do not have blackbody surface properties.
density, ρ	The amount of mass per unit volume. In heat transfer problems, the density works with the specific heat to determine how much energy a body can store per unit increase in temperature. Its units are kg/m^3.
emissive power	The heat per unit time (and per unit area) emitted by an object. For a blackbody, this is given by the Stefan-Boltzmann relation $\sigma * T^4$
graybody	A body that emits only a fraction of the thermal energy emitted by an equivalent blackbody. By definition, a graybody has a surface emissitivy less than 1, and a surface reflectivity greater than zero.
heat flux, *q*	The rate of heat flowing past a reference datum. Its units are W/m^2.
internal energy, *e*	A measure of the internal energy stored within a material per unit volume. For most heat transfer problems, this energy consists just of thermal energy. The amount of thermal energy stored in a body is manifested by its temperature.
radiation view factor, F_{12}	The fraction of thermal energy leaving the surface of object 1 and reaching the surface of object 2, determined entirely from geometrical considerations. Stated in other words, F_{12} is the fraction of object 2 visible from the surface of object 1, and ranges from zero to 1. This quantity is also known as the Radiation Shape Factor. Its units are dimensionless.
rate of heat generation, q_{gen}	A function of position that describes the rate of heat generation within a body. Typically, this new heat must be conducted to the body boundaries and removed via convection and/or radiation heat transfer. Its units are W/m^3.
specific heat, *c*	A material property that indicates the amount of energy a body stores for each degree increase in temperature, on a per unit mass basis. Its units are $J/kg - K$.
Stefan-Boltzmann constant, σ	Constant of proportionality used in radiation heat transfer, whose value is $5.669 \times 10^{-8} W/m^2 - K^4$. For a blackbody, the heat flux emitted is given by the product of σ and the absolute temperature to the fourth power.

(continued)

surface emissitivy	The relative emissive power of a body compared to that of an ideal blackbody. In other words, the fraction of thermal radiation emitted compared to the amount emitted if the body were a blackbody. By definition, a blackbody has a surface emissivity of 1. The emissivity is also equal to the absorption coefficient, or the fraction of any thermal energy incident on a body that is absorbed.
thermal conductivity, k	A material property that describes the rate at which heat flows within a body for a given temperature difference. Its units are $W/m - k$.
thermal diffusivity, α	A material property that describes the rate at which heat diffuses through a body. It is a function of the body's thermal conductivity and its specific heat. A high thermal conductivity will increase the body's thermal diffusivity, as heat will be able to conduct across the body quickly. Conversely, a high specific heat will lower the body's thermal diffusivity, since heat is preferentially stored as internal energy within the body instead of being conducted through it. Its units are m^2/s

H.1.1 Heat Transport: Conduction, Convection, Radiation

Here we will define and explain each of heat transport conditions such as Conduction, Convection and Radiation briefly.

H.1.1.1 Conduction

If different parts of an isolated solid are at different temperatures, heat will flow from the hot places to the cold ones until eventually all is at the same temperature. By "isolated" here we mean that the solid is not able to exchange heat with the outside world.

Experimentally, it is found that for most substances the rate of heat flow at any point is proportional to the temperature gradient—how fast the temperature is changing with position. To give an example, consider heat flowing down a thin rod, heated at one end, and assume the rod is wrapped in insulation so all the heat flows *down* the rod, none escapes from the surface. The natural unit of heat flow down the rod is how many joules per second pass a fixed point in the rod. It is found that:

$$\frac{dQ}{dt} \propto \frac{dT}{dx}$$

where Q is in joules, T in degrees Kelvin, x is meters down the rod. The heat flow rate is then in joules per second, or watts. It is evident from this equation that if heat is supplied at a steady rate to one end of the rod, and drains from the other end, the temperature distribution will ultimately settle down to $dT/dx =$ constant, a linear drop along the rod from one end to the other.

It is also found experimentally that a rod of double the cross-section carries twice the heat current at the same temperature difference. (This is also true for electric current, but remember it is not true for water in a pipe—the "caloric" fluid evidently doesn't act like a viscous liquid.)

This makes it possible to define a *coefficient of* κ *thermal conductivity* κ for a particular material by

$$\frac{dQ}{dt} = \kappa A \frac{dT}{dx}$$

for heat flow across an area A (in square meters) for a given temperature gradient dT/dx.

The units of κ are watts/K·meters. Some values: copper 390, stainless steel 13, glass around 0.8, white pine 0.11, air 0.026.

H.1.1.2 Microscopic Picture of Conduction

When a solid is warmed, the atoms jiggle around more—the heat energy is partially their kinetic energy of motion, partly the extra energy stores in the springy bonds between them as a result of their motion. If one end of a solid is heated, the more vigorously moving atoms there bounce against their neighbors, which then begin to move more vigorously, and the motion diffuses down the line. Obviously, this cannot be the whole story, because if we hit one end of the rod, a sound wave travels down by neighbor hitting neighbor, and moves far faster than heat. A more accurate picture (for a nonmetal) is that when one end is heated, tiny sound waves (called phonons) are generated by the fast moving atoms near the surface. These phonons travel into the solid at the speed of sound, but, unlike the massive compression wave when the end of the rod is hit, these phonons bounce off impurities or imperfections in the solid and follow random paths, only a few tens of atomic spaces between hits, typically. This, then, is very like the diffusion of a molecule in a gas we studied earlier, and it takes several minutes for heat to make its way through, say, half a centimeter of glass. The picture is different for metals: the electrons which conduct electricity so efficiently do the same for heat. However, heat transport by electrons cannot be understood without quantum mechanics: Pauli's Exclusion Principle means only about 1 % of the electrons take part in the heat conduction, but it also means that they travel far faster. The thing to remember at this stage is that the electrons carry the heat, phonons do too, but make a negligible contribution. As you might expect, good conductors of electricity are also good conductors of heat. This is why copper is used in saucepans (also, it doesn't corrode too badly).

Conductivity in liquids and gases can be measured—but usually heat transport in fluids is dominated by convection, see below. Exceptions are, for example, a fluid heated from above, or a pond below 4° Celsius being cooled from above on a winter night.

H.1.1.3 American Units

In the real world out there, the units are different. Heat flow (in construction jobs, for example) is measureds in BTU's per hour, temperature gradients in degrees Fahrenheit per inch thickness, and cross-sectional area in square feet!. The R-value of "thermal resistance" is the inverse of the thermal conductivity. For one square foot of material, one inch thick, R relates the heat current to the temperature drop by an Ohm's Law equation $\Delta T = IR$. Different areas and thicknesses scale in the obvious way. For a wall made of layers of different materials, the R-values just add.

H.1.1.4 Convection

Convection is gravitationally-induced heat transport, driven by the expansion of a fluid on heating. The hot expanded fluid has lower density, so will rise to the top of colder, and therefore denser, fluid. The simplest example is water in a kettle heated from below: hot water will rise in a central column, spread through the top layer, cooling, then flow back down around the outside. The pattern becomes more complicated if a fluid is being heated over a large area, with no obvious center. Convection cells can arise, each having a pattern like that in the kettle, the cells in a hexagonal pattern. This can happen in weather: a storm can be such a cell. However, many patterns are possible: the fluid mechanics is extremely complex. One important example of convection currents is inside the earth. Such currents deep inside are believed drive the surface movement of plates, causing earthquakes, tsunamis, etc.

H.1.1.5 Radiation

Heat from the sun reaches us as radiation, much of it visible light, the rest similar electromagnetic waves but at wavelengths our eyes are not sensitive to. All bodies not at absolute zero temperature radiate, at room temperature the radiation is in the infrared, wavelengths longer than those of the visible spectrum. Microscopically, the radiation comes about because the oscillating ions and electrons in a warm solid are accelerating electric charges, and as you will find next semester, such charges radiate. Different substances radiate with different efficiencies, those that radiate better also absorb incoming radiation better. A perfect absorber is called a black body (such perfection is not found in nature, but some things are close). This, then, is also a perfect radiator. It was found experimentally that for a perfect black body at an even temperature, the radiant energy output in watts per square meter of surface went as the fourth power of the absolute temperature:

$$P = \sigma T^4$$

P being power per square meter, σ is Stefan's constant, 5.67×10^{-8}Watts/sq.m./K^4.

For a given T, the radiant power peaks at a certain wavelength, $\lambda_{\max} T = \text{constant}$. This was well established: on heating a piece of metal, say, such as turning a dimmer on an ordinary light bulb, the first visible radiation is in the red, and an extremely hot object becomes white or even bluish. However, this was theoretically incomprehensible without quantum mechanics, and in fact the mystery of black body radiation led Planck to the first formulation of the idea of the quantum.

H.1.2 Heat Equation with Conduction and Convection

We consider the heat equation on the (0,1) with two extra terms that correspond to heat conduction and convection.

$$u_t(x,t) = k\left[\left(u_{xx}(x,t) - 2au(x,t)_x + bu(x,t)\right]\right. \quad 0 < x < L, \quad t > 0 \tag{H.1}$$

$$u(0,t) = 0 \tag{H.2}$$

$$u(a,t) \tag{H.3}$$

$$u(x,0) = \varphi(x) \tag{H.4}$$

There are many different ways to approach this problem given the boundary and initial condition and one would be able to apply separation of variable directly. The disadvantage to this is that one gets a more complicated ODE for $X(x)$ and there is a more difficult analysis of the eigenvalue and eigenvectors.

We will take a different approach which allows us to use our earlier work after a change of dependent variable. So to this end let us define another variable such as $\vartheta(x,t)$ via

$$u(x,t) = e^{ax+\beta t}\vartheta(x,t), \quad \beta = k(b - a^2) \tag{H.5}$$

Thus we have

$$\vartheta(x,t) = e^{-(ax+\beta t)}u(x,t)$$

and we can compute

$$\begin{aligned}
\vartheta_t - k\vartheta_x x &= e^{-(ax+\beta t)}(-\beta u + u_t) - k\left[\left(e^{-(ax+\beta t)}(-au + u_x)\right]_x\right. \\
&= e^{-(ax+\beta t)}\{(-\beta u + u_t) - k[-a(-au + u_x) + (-au_x + u_{xx})]\} \\
&= e^{-(ax+\beta t)}[u_t - k(u_{xx} - 2au_x + a^2 u) + \beta u] \\
&= e^{-(ax+\beta t)}[u_t - k(u_{xx} - 2au_x + a^2 u + a^2 u(b - a^2)u)] \\
&= e^{-(ax+\beta t)}[u_t - k(u_{xx} - au_x + a_2 u + bu)] = 0
\end{aligned}$$

Furthermore

$$\vartheta(0,t) = e^{-\beta t}u(0,t) = 0, \quad \vartheta(1,t) = e^{-a-\beta t}u(1,t) = 0$$

and

$$\vartheta(x,0) = e^{-ax}u(x,0) = e^{-ax}\varphi(x)$$

Therefore, $\vartheta(x,t)$ is the solution of the following?

$$\begin{aligned}
&\vartheta_t = k\vartheta_x x \\
&\vartheta(0,t) = 0 \qquad\qquad \text{and} \quad \vartheta(1,t) = 0 \\
&\vartheta(x,0) = e^{-ax}\varphi(x)
\end{aligned}$$

With $\lambda_n = -(n\pi)^2$ the solution to this problem is

$$\vartheta(x,t) = \sum_{n=1}^{\infty} b_n e^{\lambda_n t} \sin(n\pi x) \quad \text{with} \quad b_n 2\int_0^1 e^{-ax}\varphi(x)\sin(n\pi x)\,dx$$

Finally our solution to Eqs. (H.1)–(H.4) can be written as

$$u(x,t) = e^{ax+\beta t}\sum_{n=1}^{\infty} b_n e^{\lambda_n t}\sin(n\pi x)$$

H.2 Solving Heat Equation using Different Methods

For simplicity of this approach we deal with solving Heat Conduction in the Cartesian Coordinate System and we do it for one-dimensional case and then extended it to more variables (3D) and other coordinate systems. In this section we examine the solution of boundary-value problems of heat conduction in the Cartesian coordinate system for the 1D, 2D, and 3D finite, semi-finite, and infinite

regions. Separation of Variables Method will be utilized, so long as the problem falls within Homogeneous type Boundary-Value Problems. The Integral-Transform technique such as Fourier, Laplace and other transformation will be applied for the solution of nonhomogeneous problems.

As we said the method of separation of variables has been used in the solution of homogeneous heat conduction problem. The multidimensional steady-state heat conduction problems with no heat flux or generation can also be solved with this method if only one of the boundary conditions is nonhomogeneous; Problem involving more than one non-homogeneous boundary conditions can be split up into simpler problems each containing only one nonhomogeneous boundary condition

H.2.1 Separation of Variables Method

The method of Separation of Variables cannot always be used and even when it can be used it will not always be possible to get much past the first step in the method. However, it can be used to easily solve the 1D heat equation with no sources, the 1D wave equation, and the 2D version of Laplace's Equation, $\nabla^2 T = 0$.

In order to use the method of separation of variables we must be working with a linear homogenous partial differential equations with linear homogeneous boundary conditions. At this point we're not going to worry about the initial condition (s) because the solution that we initially get will rarely satisfy the initial condition (s). As we'll see however there are ways to generate a solution that will satisfy initial condition(s) provided they meets some fairly simple requirements.

The method of separation of variables relies upon the assumption that a function of the form,

$$T(x,t) = X(x)Y(t) \tag{H.6}$$

will be a solution to a linear homogeneous partial differential equation in x and t. This is called a product solution and provided the boundary conditions are also linear and homogeneous this will also satisfy the boundary conditions. However, as noted above this will only rarely satisfy the initial condition, but that is something for us to worry about in the next section.

Now, before we get started on some examples there is probably a question that we should ask at this point and that is: Why?. Why did we choose this solution and how do we know that it will work? This seems like a very strange assumption to make. After all there really isn't any reason to believe that a solution to a partial differential equation will in fact be a product of a function of only x's and a function of only t's. This seems more like a hope than a good assumption/guess.

Unfortunately the best answer is that we chose it because it will work. As we'll see it works because it will reduce our partial differential equation down to two ordinary differential equations and provided we can solve those then we're in

business and the method will allow us to get a solution to the partial differential equations [1].

So, let's do a couple of examples to see how this method will reduce a partial differential equation down to two ordinary differential equations.

Example 1:
Use Separation of Variables on the following Partial Differential Equation.

$$\frac{\partial T(x,T)}{\partial t} = k\frac{\partial^2 T(x,t)}{\partial x^2}$$

$$\text{Initial Condition}\,\mathbf{I.C.}\,T(x,0) = f(x)$$

$$\text{Boundary Condition.}\,\mathbf{B.C.}\quad \begin{aligned} T(0,t) &= 0 \\ T(L,t) &= 0 \end{aligned}$$

Solution:
So, we have the heat equation with no sources, fixed temperature boundary conditions (that are also homogeneous) and an initial condition. The initial condition is only here because it belongs here, but we will be ignoring it until we get to the next section.

The method of separation of variables tells us to assume that the solution will take the form of the product,

$$T(x,t) = X(x)Y(t)$$

so all we really need to do here is plug this into the differential equation and see what we get.

$$\frac{\partial}{\partial t}(X(x)Y(t)) = k\frac{\partial^2}{\partial x^2}(X(x)Y(t))$$

$$X(x)\frac{\partial Y(t)}{\partial t} = kY(t)\frac{\partial^2 X(x)}{\partial x^2}$$

As shown above we can factor the $X(x)$ out of the time derivative and we can factor the $Y(t)$ out of the spatial derivative. Also notice that after we've factored these out we no longer have a partial derivative left in the problem. In the time derivative we are now differentiating only $Y(t)$ with respect to t

(continued)

and this is now an ordinary derivative. Likewise, in the spatial derivative we are now only differentiating $X(x)$ with respect to x and so we again have an ordinary derivative.

At this point it probably doesn't seem like we've done much to simplify the problem. However, just the fact that we've gotten the partial derivatives down to ordinary derivatives is liable to be good thing even if it still looks like we've got a mess to deal with.

Speaking of that apparent (and yes I said apparent) mess, is it really the mess that it looks like?. The idea is to eventually get all the t's on one side of the equation and all the x's on the other side. In other words we want to "separate the variables" and hence the name of the method. In this case let's notice that if we divide both sides by $X(x)Y(t)$ we get want we want and we should point out that it won't always be as easy as just dividing by the product solution. So, dividing out gives us,

$$\frac{1}{Y}\frac{dY}{dt} = k\frac{1}{X}\frac{d^2X}{dx^2} \quad \Rightarrow \quad \frac{1}{kY}\frac{dY}{dt} = \frac{1}{X}\frac{d^2X}{dx^2}$$

Notice that we also divided both sides by k. This was done only for convenience down the road. It doesn't have to be done and nicely enough if it turns out to be a bad idea we can always come back to this step and put it back on the right side. Likewise, if we don't do it and it turns out to maybe not be such a bad thing we can always come back and divide it out. For the time being however, please accept our word that this was a good thing to do for this problem. We will discuss the reasoning for this after we're done with this example.

Now, while we said that this is what we wanted it still seems like we've got a mess. Notice however that the left side is a function of only t and the right side is a function only of x as we wanted. Also notice these two functions must be equal.

Let's think about this for a minute. How is it possible that a function of only t's can be equal to a function of only x's regardless of the choice of t and/or x that we have?. This may seem like impossibility until you realize that there is one way that this can be true. If both functions (i.e., both sides of the equation) were in fact constant and not only a constant, but the same constant then they can in fact be equal.

So we must have:

$$\frac{1}{kY}\frac{dY}{dt} = \frac{1}{X}\frac{d^2X}{dx^2} = -\lambda$$

where the $-\lambda$ is called the **Separation Constant** and is arbitrary.

(continued)

The next question that we should now address is why the minus sign?. Again, much like the dividing out the k above, the answer is because it will be convenient down the road to have chosen this. The minus sign doesn't have to be there and in fact there are times when we don't want it there.

So how do we know it should be there or not?. The answer to that is to proceed to the next step in the process (which we'll see in the next section) and at that point we'll know if would be convenient to have it or not and we can come back to this step and add it in or take it our depending what we chose to do here.

Okay, let's proceed with the process. The next step is to acknowledge that we can take the equation above and split it into the following two ordinary differential equations.

$$\frac{dY}{dt} = -k\lambda Y \quad \frac{d^2X}{dx^2} = -\lambda X$$

Both of these are very simple differential equations; however, because we don't know what λ is, we actually can't solve the spatial one yet. The time equation however could be solved at this point if we wanted to, although that won't always be the case. At this point we don't want to actually think about solving either of these yet however.

The last step in the process that we'll be doing in this section is to also make sure that our product solution, $T(x,t) = X(x)Y(t)$, satisfies the boundary conditions, so let's plug it into both equations of those.

$$T(0,t) = X(0)Y(t) = 0 \quad T(L,t) = X(L)Y(t) = 0$$

Let's consider the first one for a second. We have two options here. Either $X(0) = 0$ or $Y(t) = 0$ for every t. However, if we have $Y(t) = 0$ for every t then we'll also have $T(x,t) = 0$, *i.e.* the trivial solution, and as we discussed in the previous section this is definitely a solution to any linear homogeneous equation we would really like a non-trivial solution.

Therefore we will assume that in fact we must have $X(0) = 0$. Likewise, from the second boundary condition we will get $Y(t) = 0$ to avoid the trivial solution. Note as well that we were only able to reduce the boundary conditions down like this because they were homogeneous. Had they not been homogeneous we could not have done this.

So, after applying separation of variables to the given partial differential equation we arrive at a 1st order differential equation that we'll need to solve for and a 2nd order boundary value problem that we'll need to solve for $X(x)$. The point of this section however is just to get to this point and we'll hold off solving these until the next section.

(continued)

Let's summarize everything up that we've determined here.

$$\frac{dY}{dt} = -k\lambda Y \quad \frac{d^2X}{dx^2} + \lambda X = 0$$

$$X(0) = 0 \quad X(L) = 0$$

and note that we don't have a condition for the time differential equation and is not a problem. Also note that we rewrote the second one a little.

The time dependent equation can really be solved at any time, but since we do not know what λ is yet let's hold off on that one. Also note that in many problems only the boundary value problem can be solved at this point so do not always expect to be able to solve either one at this point.

The spatial equation is a Boundary Value Problem (BVP) and we know from our work in Sect. 1.10 Appendix C that it will only have non-trivial solution (which we want) for certain values of λ, which we will recall are called eigenvalue. Once we have those we can determine the non-trivial solutions for each λ, *i.e.* eigenfunctions.

Now, we actually solved the spatial problem,

$$\frac{d^2X}{dx^2} + \lambda X = 0$$

$$X(0) = 0 \quad X(L) = 0$$

In Example 1 of the Sect. 3.0 of Appendix C for $L = 2\pi$. So, because we have solved this once for a specific L and the work is not all that much different for a general L we are not going to put a lot of effort here and if you need to review just go to Example 1 from Sect. 3.0 of Appendix C.

We have three cases to deal with as follows:

$$\underline{\lambda > 0}$$

In this case we know the solution to the differential equation is,

$$X(x) = c_1 \cos\left(\sqrt{\lambda x}\right) + c_2 \sin\left(\sqrt{\lambda x}\right)$$

Applying again the first boundary condition gives,

$$0 = X(0) = c_1$$

Now applying the second boundary condition, and using the above result of course, gives,

(continued)

$$0 = X(L) = c_2 \sin\left(L\sqrt{\lambda}\right)$$

Now, we are after nontrivial solutions and this means we must have:

$$\sin\left(L\sqrt{\lambda}\right) = 0 \quad \Rightarrow \quad X_n(x) = \sin\left(\frac{n\pi x}{L}\right) \quad n = 1, 2, 3, \ldots$$

Note that we did not need the c_2 in the eigenfunction as it will just get absorbed into another constant that we will picking up later on.

$$\underline{\lambda > 0}$$

The solution to the differential equation in this case is

$$X(x) = c_1 + c_2 x$$

Applying the boundary conditions gives,

$$0 = X(0) = c_1 \quad 0 = X(L) = c_2 L \quad c_2 = 0$$

So, in this case the only solution is the trivial solution and so $\lambda = 0$ is not an eigenvalue for this boundary value problem.

$$\underline{\lambda > 0}$$

Here the solution to the differential equation is:

$$X(x) = c_1 \cosh\left(\sqrt{-\lambda}x\right) + c_2 \sinh\left(\sqrt{-\lambda}x\right)$$

Applying the first boundary condition gives,

$$0 = X(0) = c_1$$

and applying the second gives,

$$0 = X(L) = c_2 \sinh\left(L\sqrt{-\lambda}\right)$$

So, we are assuming $\lambda < 0$ and so $L\sqrt{-\lambda} \neq 0$ and this means $\sinh\left(L\sqrt{-\lambda}\right) \neq 0$. We therefore we must have and so we can only get the trivial solution in this case.

(continued)

Therefore, there will be no negative eigenvalues for this boundary value problem. The complete list of eigenvalues and eigenfunctions for this problem are then,

$$\lambda_n = \left(\frac{n\pi}{L}\right)^2 \quad X_n(x) = \sin\left(\frac{n\pi x}{L}\right) \quad n = 1, 2, 3, \ldots$$

Now let us solve the time differential equation,

$$\frac{dY}{dt} = k\lambda_n Y$$

And note that even though we know λ we are not going to plug it in quite yet to keep the mess to a minimum. We will however now use λ_n to remind us that we actually have an infinite number of possible values here. This is a simple linear (and separable for that matter) 1st order differential equation and so we will let you verify that the solution is:

$$Y(t) = ce^{-k\lambda_n t} = ce^{-k\left(\frac{n\pi}{L}\right)^2 t}$$

Now that we have both ODEs solved we can finally write down a solution. Note however that we have in fact found infinitely many solutions since there are infinitely many solutions (*i.e.* eigenfunctions) to the spatial problem. Our solution is a product of both solutions are then

$$u_n(x,t) = B_n \sin\left(\frac{n\pi x}{L}\right) e^{-k\left(\frac{n\pi}{L}\right)^2 t} \quad n = 1, 2, 3, \ldots$$

We've denoted the product solution u_n to acknowledge that each value of n will yield a different solution. Also note that we've changed the c in the solution to the time problem to B_n and to denote the fact that it will probably be different for each value of n as well and because had we kept the c_2 with the eigenfunction we'd have absorbed it into the c to get a single constant in our solution. The function above will satisfy the heat equation and the boundary condition of zero temperature on the ends of the bar.

Okay, so just what have we learned here?. By using separation of variables we were able to reduce our linear homogeneous partial differential equation with linear homogeneous boundary conditions down to an ordinary differential equation for one of the functions in our product solution (Eq. H.6), $Y(t)$ in this case, and a boundary value problem that we can solve for the other function, $X(x)$ in this case.

Note as well that the boundary value problem is in fact an eigenvalue/eigenfunction problem. When we solve the boundary value problem we will be identifying the eigenvalues, λ, that will generate non-trivial solutions the their corresponding eigenfunctions. Again, we'll look into this more in the next section. At this point all we want to do is identify the two ordinary differential equations that we need to solve to get a solution.

Before we do a couple of other examples we should take a second to address the fact that we made two very arbitrary seeming decisions in the above work. We divided both sides of the equation by k at one point and chose to use $-\lambda$ instead of λ as the separation constant.

Both of these decisions were made to simplify the solution to the boundary value problem we got from our work. The addition of the k in the boundary value problem would just have complicated the solution process with another letter we'd have to keep track of so we moved it into the time problem were it won't cause as many problems in the solution process. Likewise, we chose $-\lambda$ because we've already solved that particular boundary value problem (albeit with a specific L, but the work will be nearly identical) when we first looked at finding eigenvalues and eigenfunctions. This by the way was the reason we rewrote the boundary value problem to make it a little clearer that we have in fact solved this one already.

We can now at least partially answer the question of how do we know to make these decisions. We wait until we get the ordinary differential equations and then look at them and decide of moving things like the k or which separation constant to use based on how it will affect the solution of the ordinary differential equations. There is also, of course, a fair amount of experience that comes into play at this stage. The more experience you have in solving these easier it often is to make these decisions.

Again, we need to make clear here that we're not going to go any farther in this section than getting things down to the two ordinary differential equations. Of course we will need to solve them in order to get a solution to the partial differential equation but that is the topic of the remaining sections in this chapter. All we'll say about it here is that we will need first to solve the boundary value problem, which will tell us what λ must be and then we can solve the other differential equation. Once that is done we can then turn our attention to the initial condition.

Okay, we need to work a couple of other examples and these will go a lot quicker because we won't need to put in all the explanations. After the first example this process always seems like a very long process but it really isn't. It just looked that way because of all the explanation that we had to put into it.

Example 2:
Use Separation of Variable on the following partial differential equation

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial t^2}$$

$$u(x,0) = f(x) \quad \frac{\partial u(0,t)}{\partial x} = 0 \quad \frac{\partial u(L,t)}{\partial x} = 0$$

Solution:
In this case we are looking at the heat equation with no sources and perfectly insulated boundaries

So, we will start off by again assuming that our product solution will have the form

$$u(x,t) = \varphi(x)T(t)$$

and because the differential itself has not changed here we will get the same result from plugging this in as we did in the previous example so the two ordinary differential equation that we will need to solve are:

$$\frac{dT}{dt} = -k\lambda T \quad \frac{d^2\varphi}{dx^2} = -\lambda\varphi$$

Now, the point of this example was really to deal with the boundary conditions so let us plug the product solution into them to get,

$$\frac{\partial(T(t)\varphi(x))(0,t)}{\partial x} = 0 \quad \frac{\partial(T(t)\varphi(x))(L,t)}{\partial x} = 0$$

$$T(t)\frac{d\varphi(0)}{dx} = 0 \qquad T(t)\frac{d\varphi(L)}{dx} = 0$$

Now, just as with the first example if we want to avoid the trivial solution and so we cannot have $T(t) = 0$ for every t and so we must have

$$\frac{d\varphi(0)}{dx} = 0 \quad \frac{d\varphi(L)}{dx} = 0$$

Here is a summary of what we get by applying separation of variables to this problem.

(continued)

$$\frac{dT}{dt} = -k\lambda T \qquad \frac{d^2\varphi}{dx^2} + \lambda\varphi = 0$$
$$\frac{d\varphi(0)}{dx} = 0 \quad \frac{d\varphi(L)}{dx} = 0$$

Next, let us see what we get if use periodic boundary conditions with the heat equation.

Example 3:
Use Separation of Variable on the following partial differential equation

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial t^2}$$
$$u(x,0) = f(x) \quad u(-L,t) = u(L,t) \quad \frac{\partial u(-L,t)}{\partial x} = \frac{\partial u(L,t)}{\partial x}$$

Solution:
First note that these boundary conditions really are homogeneous boundary conditions. If we rewrite them as,

$$u(-L,t) - u(L,t) = 0 \quad \frac{\partial u(-L,t)}{\partial x} - \frac{\partial u(L,t)}{\partial x} = 0$$

It is a little easier to see.

Now, again we have done this partial differential equation so we will start off with,

$$u(x,t) = \varphi(x)T(t)$$

and the two ordinary differential equations that we will need to solve are:

$$\frac{dT}{dt} = -k\lambda T \qquad \frac{d^2\varphi}{dx^2} = -\lambda\varphi$$

Plugging the product solution into the rewritten boundary conditions gives,

(continued)

$$T(t)\varphi(-L) - T(t)\varphi(L) = T(t)[\varphi(-L) - \varphi(L)] = 0$$

$$T(t)\frac{\varphi(-L)}{dx} - T(t)\frac{\varphi(L)}{dx} = T(t)\left[\frac{d\varphi(-L)}{dx} - \frac{d\varphi(L)}{dx}\right] = 0$$

And we can see that we will only get non-trivial solution if,

$$\varphi(-L) - \varphi(L) = 0 \quad \frac{d\varphi(-L)}{dx} - \frac{d\varphi(L)}{dx} = 0$$

$$\varphi(-L) = \varphi(L) \qquad \frac{d\varphi(-L)}{dx} = \frac{d\varphi(L)}{dx}$$

So, here is what we get by applying separation of variables to this problem

$$\frac{dT}{dt} = -k\lambda T \qquad \frac{d^2\varphi}{dx^2} + \lambda\varphi = 0$$

$$\varphi(-L) = \varphi(L) \quad \frac{\varphi(-L)}{dx} = \frac{\varphi(L)}{dx}$$

http://tutorial.math.lamar.edu/Classes/DE/SolvingHeatEquation.aspx#PDE_HeatEqn_Ex1 [2]

H.2.2 Solving Heat Equation using Fourier Transform Method

The heat equation is a partial differential equation. Prior to Fourier's work, there was no known solution to the heat equation in a general situation, although particular solutions were known if the heat source behaved in a simple way, in particular, if the heat source was a sine or cosine wave. These simple solutions are now sometimes called eigensolutions. Fourier's idea was to model a complicated heat source as a superposition (or linear combination) of simple sine and cosine waves, and to write the solution as a superposition of the corresponding eigensolutions. This superposition or linear combination is called the Fourier series.

Although the original motivation was to solve the heat equation, it later became obvious that the same techniques could be applied to a wide array of mathematical and physical problems

Example 1: Use finite Fourier transforms to solve

$$\frac{\partial^2 u(x,t)}{\partial x^2} = \frac{\partial u(x,t)}{\partial t}$$

$$\textbf{B.C.} \quad \begin{cases} U(o,t) = 0 \\ U(4,t) = 0 \end{cases}$$

$$\textbf{I.C.} \quad \{U(x,0) = 2x \quad \text{where } 0 < x < 4 \text{ and } t > 0$$

Solution: Take the finite Fourier Sine Transform (with $L = 4$) (See Appendix E and Sect. 1.15) of both sides of the partial differential equation to obtain

$$\int_0^4 \frac{\partial^2 U(x,t)}{\partial x^2} \sin \frac{n\pi x}{4} dx = \int_0^4 \frac{\partial U(x,t)}{\partial t} \sin \frac{n\pi x}{4} dx$$

Writing $u = \mathbb{F}_{\text{sine}}\{T\}$ and using [See Appendix E and Sect. 1.21 Example 2(a)] with the conditions $U(0,t) = 0$ and $U(4,t) = 0$, we find

$$\frac{du}{dt} = -\frac{n^2\pi^2}{16} u \tag{H.7}$$

where $u = u(n,t)$.

Taking the finite Fourier sine transform of the condition $T(x,0) = 2x$, we have as in Example 2 (a), Sect. 1.12 of Appendix E

$$\begin{aligned} u(n,0) = f_{\text{sine}}(2x) &= \int_0^4 2x \sin\left(\frac{n\pi x}{4}\right) dx \\ &= \left\{(2x)\left(\frac{-\cos(n\pi x/4)}{n\pi/4}\right) - 2\left(\frac{-\sin(n\pi x/4)}{n^2\pi^2/16}\right)\right\}\Big|_0^4 \\ &= \frac{32(1 - \cos(n\pi))}{n\pi} \end{aligned} \tag{H.8}$$

Solving the differential Eq. (H.8), we find if c is an arbitrary constant, then we have

$$u = u(n,t) = ce^{-n^2\pi^2 t/16} \tag{H.9}$$

But if use the initial condition for $t = 0$, then $c = u(n,0)$ which from (H.11) and (H.12) results;

$$u = \frac{32(1 - \cos(n\pi))}{n\pi} e^{-(n^2\pi^2)t/16}$$

Thus from Example 1 (a) Sect. 1.12 of Appendix E, the inverse Fourier sine transform is

$$U(x,t) = \frac{2}{4}\sum_{n=1}^{\infty} \frac{32(1 - \cos(n\pi))}{n\pi} e^{-(n^2\pi^2)t/16}$$

$$= \frac{2}{4}\sum_{n=1}^{\infty} \frac{(1 - \cos(n\pi))}{n} e^{-(n^2\pi^2)t/16}$$

Physically $U(x,t)$ represents the temperature at any point x at any time t in a solid bounded by the point $x = 0$ and $x = 4$. The conditions $u(0,t) = 0$ and $u(4,t) = 0$ express the fact that the ends are kept at temperature zero, while $u(x,0) = 2x$ express the initial temperature as a function of x. Equivalently, the solid can be replaced by a bar on the axis with endpoint at $x = 0$ and $x = 4$ whose surface is insulated.

Example 4: Solve $\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}$ for $x > 0$, $t > 0$ subject to the condition

$$U(0,t) = 0 \qquad U(x,0) = \begin{cases} 1 & 0 < x < 1 \\ 0 & x \geq 1 \end{cases} \qquad U(x,t) \text{ is bounded}$$

Taking the Fourier sine transform of both sides of the given partial differential equation, we find

$$\int_0^\infty \frac{\partial U(x,t)}{\partial t} \sin(\lambda x)\, dx = \int_0^\infty \frac{\partial^2 U}{\partial x^2} \sin(\lambda x) dx \tag{H.10}$$

Then if $u = u(\lambda, t) = \int_0^\infty U(x,t) \sin(\lambda x) dx$

$$\text{This becomes } \frac{du}{dt} = \left\{ \frac{\partial U(x,t)}{\partial x} \sin(\lambda x) - \lambda U(x,t) \cos(\lambda x) \right\} \Big|_0^\infty$$
$$-\lambda^2 \int_0^\infty U(x,t) \sin(\lambda x) dx = \lambda U(0,t) - \lambda^2 u \tag{H.11}$$

on integrating the right hand side of Eq. (H.10) by parts and assuming that $u(x,t)$ and $U(x,t)/\partial x$ approach zero as $x \to \infty$.

From the condition for $U(x,0)$, we have on taking the Fourier sine transform

$$U(\lambda, 0) = \int_0^\infty U(x,0) \sin(\lambda x) dx$$
$$= \int_0^1 \sin(\lambda x) dx = \frac{1 - \cos\lambda}{\lambda} \tag{H.12}$$

Solving Eq. (H.11) subject to the condition Eq. (H.12) and $U(0,t) = 0$, we find

$$u(\lambda, t) = \frac{1 - \cos\lambda}{\lambda} e - \lambda^2 t$$

Then taking the inverse Fourier sine transform, we find the required solution

$$u(\lambda, t) = \frac{2}{\pi} \int_0^1 \frac{1 - \cos\lambda}{\lambda} e^{-\lambda^2 t} \sin(\lambda x) dx d\lambda$$

Physically, this can represent the temperature in a solid $x > 0$. See Example 1 above

H.2.3 Solving Heat Equation using Laplace Transform Method

Partial differential equations, like their one-variable counterpart, ordinary differential equations, are ubiquitous throughout the scientific spectrum. However, they are, in general, more difficult to

solve. Yet here again, we may apply the Laplace transform method to solve PDEs by reducing the initial problem to a simpler ODE.

Partial differential equations come in three types. For a function of two variables $u = u(x,t)$ where we have replace T with u, the general second-order linear PDE has the form Equation C.75 of Sect. 6.1 of Appendix C and in this section we are much interested in parabolic aspect of PDE which is in the form of one-dimensional *Heat Equation* as follows [3];

$$\frac{\partial u(x,t)}{\partial t} = c \frac{\partial^2 u(x,t)}{\partial x^2} \text{ (parabolic form)} \tag{H.13}$$

In its simple form, we have assumed the above partial differential equation is function of one variable x and it is a linear class.

Laplace Transform Method is thoroughly explained in Sect. 2.0 Appendix E and we remind of it here in a very brief steps. We consider the function $u = u(x,t)$, where $t \geq 0$ is a time variable. Denote by $U(x,s)$ the *Laplace transform of u with respect to t*, that is, to say

$$U(x,s) = \pounds\{u(x,t)\} = \int_0^\infty e^{-st} u(x,t) dt$$

Here x is the "gun transformed variable". There are few examples that are given in Appendix E under Sect. 2 that readers should refer to them for more familiarity. Here we provide few example given by J. Schiff [3].

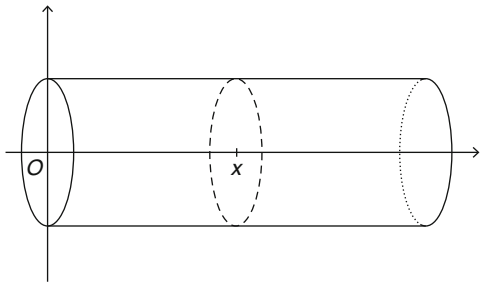


Fig. H.1 Finite or semi-infinite thin rod

The heat flow in a finite or semi-infinite thin rod is governed by the PDE

$$\frac{\partial u(x,t)}{\partial t} = c\frac{\partial^2 u(x,t)}{\partial x^2}$$

where c is a constant (called the *diffusitivity*), and $u(x,t)$ is the temperature at position x and time t over cross-section of this rod. The temperature over this cross-section at is taken to be uniform (See Fig. H.1).

Example 5: Solve

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2} \quad x > 0 \quad \text{and} \quad t > 0 \text{ Heat PDE} \tag{H.14}$$

For Boundary and Initial conditions as follows:

$$\{(\text{i}) \;\; u(x,0) = 1 \quad \text{for } x > 0 \quad \text{Intitial Condition (I. C.)}$$

$$\begin{cases} (\text{ii}) & \dfrac{\partial u(0,t)}{\partial x} = 0 \text{ for } t > 0 \\ (\text{iii}) & \lim\limits_{x\to\infty} u(x,t) = 1 \end{cases} \quad \text{Boundary Condition (B. C.)}$$

Solution: Take Laplace Transform of above heat PDE and Eq. (H.14) yields

$$\frac{d^2U(x,s)}{dx^2} = sU(x,s) - u(x,0) = sU(x,s) - 1 \tag{H.15}$$

Laplace transform of B.C. (ii) and (iii) provides

$$U(0,s) = \pounds\{u(0,t)\} = 0$$

$$\lim_{x\to\infty} U(x,s) = \lim_{x\to\infty} \pounds\{u(x,t)\} = \pounds\left\{\lim_{x\to\infty} u(x,t)\right\} = \frac{1}{s}$$

Now Eq. (H.15) is just and Ordinary Differential Equation (ODE) whose solution is given by:

$$U(x,s) = c_1 e^{\sqrt{sx}} + c_2 e^{-\sqrt{sx}} + \frac{1}{s}$$

The boundary condition $\lim_{x\to\infty} U(x,s) = 1/s$ implies $c_1 = 0$ and $U(0,s) = 0$ implies

$$U(x,s) = \frac{1}{s} - \frac{e^{\sqrt{sx}}}{s}$$

By Example 1 Sect. 2.10 of Appendix E we have

$$u(x,t) = erf\left(\frac{x}{2\sqrt{t}}\right) = \frac{2}{\sqrt{\pi}}\int_0^{x/2\sqrt{t}} e^{-u^2} du$$

Direct calculation shows that indeed satisfies Eq. (H.14) above and that the initial and boundary conditions are satisfied.

Example 6: Solve

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2} \quad \text{in domain} \quad x > 0 \text{ and } t > 0$$

For

$$\begin{cases} \text{(i)} & u(x,0) = f(x) \qquad \text{I.C.} \\ \text{(ii)} & u(0,t) = 0 \text{ for } t > 0 \\ \text{(iii)} & u(1,t) = 0 \text{ for } t > 0 \end{cases}$$

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = -f(x)$$

$$Y(x) = U(x,s)$$

$$Y(0) = U(0,s) = 0$$

Solution: The transformed equation is

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = 0$$

Whose solution is given by

$$U(x,s) = c_2 e^{-\sqrt{sx}}$$

in view of condition (iii). By (ii)

$$U(0,s) = \pounds\{f(t)\} = F(s)$$

so that $c_2 = F(s)$ and

$$U(x,s) = F(s)e^{-\sqrt{sx}}$$

Invoking the fact that $\pounds^{-1}\left\{e^{-a\sqrt{s}}\right\} = \frac{a}{2\sqrt{\pi t^3}} e^{-a/4t}$ for $a > 0$ and the convolution theorem, we have the following results:

$$u(x,t) = \int_0^t \frac{x}{2\sqrt{\pi\tau^3}} e^{-x^2/4\tau} f(t-\tau)d\tau$$

Making the substitution $\sigma^2 = x^2/4\tau$ then we have:

$$u(x,t) = \frac{2}{\sqrt{\pi}} \int_{x/2\sqrt{t}}^{\infty} \frac{x}{2\sqrt{\pi\tau^3}} e^{-\sigma^2} f\left(t - \frac{x^2}{4\sigma^2}\right) d\sigma$$

Which is the desired solution.

Example 7 Solve

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2} \text{ in domain } 0 < x < 1 \text{ and } t > 0$$

For

$$\{\text{(i) } u(x,0) = f(x) \text{ I.C.}$$

$$\begin{cases} \text{(ii)} & u(0,t) = 0 \text{ for } t > 0 \\ \text{(iii)} & u(1,t) = 0 \text{ for } t > 0 \end{cases}$$

Solution Therefore,

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = -f(x)$$

Here we solve this ODE by the Laplace transform method as well. To this end, let $Y(x) = U(x,s)$. Then $Y(0) = U(0,s) = 0$, $Y(1) = U(1,s) = 0$. Setting $a^2 = s$, we obtain

$$\sigma^2 \pounds(Y) - \sigma Y(0) - Y'(0) - a^2\sigma(Y) = -\pounds(f) = -F(\sigma)$$

that is

$$\pounds(Y) = \frac{Y'(0)}{\sigma^2 - a^2} - \frac{F(\sigma)}{\sigma^2 - a^2}$$

Inverting gives

$$Y(x) \qquad U(x,s) = \frac{Y'(0)\sinh ax}{a} - \frac{1}{a}\int_0^x f(u)\sin ga(x-u)du$$

$$= \frac{Y'(0)\sinh\sqrt{sx}}{\sqrt{s}} - \frac{1}{\sqrt{s}}\int_0^x f(u)\sinh\sqrt{s}(x-u)du$$

Thus,

$$U(x,s) = \int_0^1 f(u)\frac{\sinh\sqrt{sx\sinh\sqrt{s(1-u)}}}{\sqrt{s}\sinh\sqrt{s}}du$$

$$-\int_0^x f(u)\frac{\sinh\sqrt{s(x-u)}}{\sqrt{s}}du$$

We can write $\int_0^1 = \int_0^x + \int_x^1$ and use the fact that $\sinh(z \pm w)$ $\sinh z\cosh w \pm \cosh z\sinh w$

Then

$$U(x,s)=\int_0^x f(u)\left[\frac{\sinh\sqrt{sx\sinh\sqrt{(1-u)}}}{\sqrt{s}\sinh\sqrt{s}}-\frac{\sinh\sqrt{s(1-u)}}{\sqrt{s}}\right]du$$

$$+\int_0^x f(u)\frac{\sinh\sqrt{s}x\sinh\sqrt{s(1-u)}}{\sqrt{s}\sinh\sqrt{s}}du$$

$$=\int_0^x f(u)\frac{\sinh\sqrt{s}(1-x)\sinh\sqrt{su}}{\sqrt{s}\sinh\sqrt{s}}du$$

$$+\int_0^1 f(u)\frac{\sinh\sqrt{s}x\sinh\sqrt{s}(1-u)}{\sqrt{s}\sinh\sqrt{s}}du$$

To determine the inverse we use the complex inversion formula (See Appendix D Sect. 30.3). When it is applied to the first integral, we have:

$$\frac{1}{2\pi i}\int_{x_0-i00}^{x_0+i00}e^{ts}\left\{\int_0^x f(u)\frac{\sinh\sqrt{s(1-x)\sinh\sqrt{su}}}{\sqrt{s}\sinh\sqrt{s}}du\right\}ds=\sum\text{Res}$$

There are simple poles in this case at $s_0=0$ and $s_n=-n^2\pi^2\ \ n-1,2,3,\cdots$

$$\text{Res}(0)=\lim_{s\to 0}s\int_0^x f(u)\frac{\sinh\sqrt{s}(1-x)\sinh\sqrt{su}}{\sqrt{s}\sinh\sqrt{s}}\,du=0$$

$$\text{Res}(-n^2\pi^2)$$

$$=\lim_{s\to -n^2\pi^2}\left(s+n^2\pi^2\right)e^{ts}\int_0^x f(u)\frac{\sinh\sqrt{s}(1-x)\sinh\sqrt{su}}{\sqrt{s}\,\sinh\sqrt{s}}du$$

$$=\lim_{s\to -n^2\pi^2}\frac{s+n^2\pi^2}{\sinh\sqrt{s}}\lim_{s\to -n^2\pi^2}e^{ts}\int_0^x f(u)\frac{\sinh\sqrt{s}(1-x)\sinh\sqrt{su}}{\sqrt{s}\,\sinh\sqrt{s}}du$$

$$=2e^{-n^2\pi^2t}\int_0^x f(u)\frac{\sinh[(n\pi i)(1-x)]\sinh(n\pi i)u}{\cosh(n\pi i)}du$$

$$=2e^{-n^2\pi^2t}\int_0^x f(u)\frac{\sinh[n\pi(1-x)]\,\sin n\pi u}{-\cos n\pi}du$$

Where we have used the properties from Appendix D with $z=x+iy$

$$\sinh z = \cos y \sinh y + i \sin y \cosh x$$

$$\cosh z = \cos y \cosh x + i \sin y \sinh x$$

To obtain the last equality. Therefore, we have

$$\sum \text{Res} = 2\sum_{n-1}^{\infty} e^{-n^2\pi^2 t}\left(\int_0^x f(u)\sin n\pi u du\right)\sin n\pi x$$

Similarly, the inverse of the second integral is given by:

$$2\sum_{n=1}^{\infty} e^{-n^2\pi^2 t}\left(\int_0^1 f(u)\sin n\pi u du\right)\sin n\pi x$$

Finally

$$u(x,t) = 2\sum_{n=1}^{\infty} e^{-n^2\pi^2 t}\left(\int_0^1 f(u)\sin n\pi u du\right)\sin n\pi x$$

The same result is obtained when we solve this example by using separation of variables method.

H.3 References

1. http://tutorial.math.lamar.edu/Classes/DE/DE.aspx
2. http://tutorial.math.lamar.edu/Classes/DE/SolvingHeatEquation.aspx
3. Schiff JL (1991) The laplace transform, theory and applications. Springer

Appendix I
Data and Plots of Thermal Parameters of Different Materials

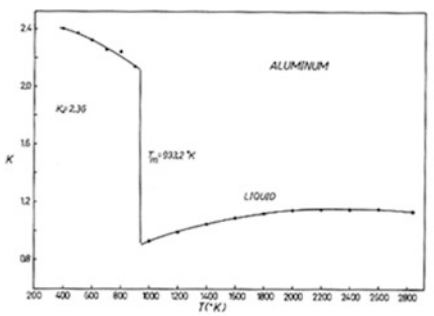
One of the major goals of physics and so is ours in case of laser in particular and its interaction with materials is to understand the nature of light. Due to the complexity nature of the light this goal is very difficult to fully achieve, but this complication means that light offers many opportunities for different applications including optical interferences and in our case response of materials to laser radiation and its interaction with matter and specifically metallic materials.

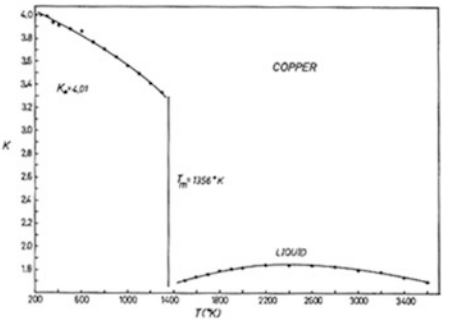
I.1 Thermal Conductivity Data

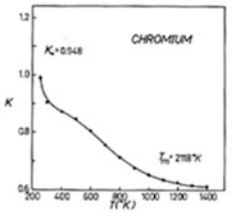
Plots of K versus T are given for aluminum, copper, chromium, cobalt, gold, iron, lead, molybdenum, nickel, platinum, rhodium, silver, tantalum, tin, titanium, tungsten, uranium, vanadium, zinc, zirconium, Armco iron, 302, 303, and 304 stainless steel, aluminum oxide, fused quartz, magnesium oxide, and titanium dioxide. K is in units of W/cm °C.

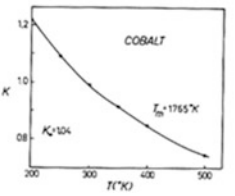
© Springer International Publishing Switzerland 2016

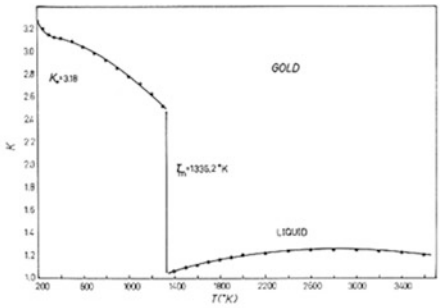
B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

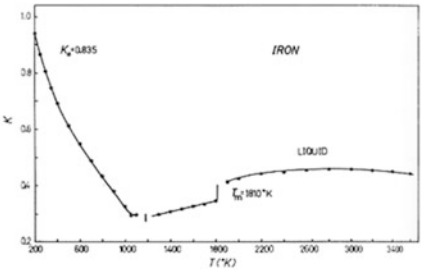

ALUMINUM
K
T(°K)
LIQUID

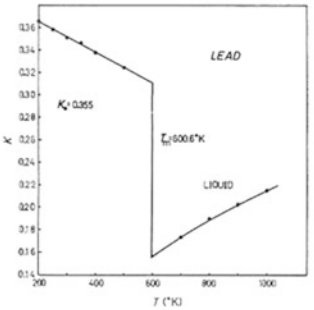

COPPER
K
T(°K)
LIQUID

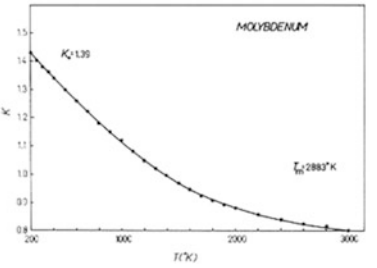

CHROMIUM
K
T(°K)

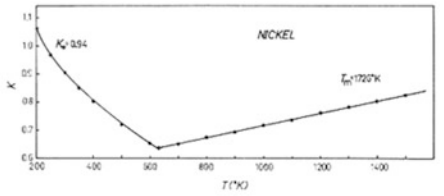

COBALT
K
T(°K)

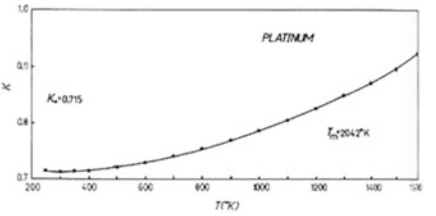

GOLD
K_s=3.18
T_m=1336.2 °K
LIQUID
K
T(°K)

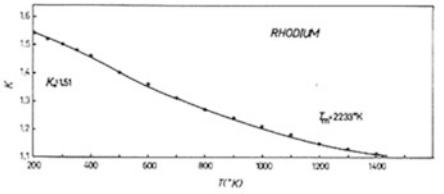

IRON
K_s=0.835
LIQUID
T_m=1810 °K
K
T(°K)

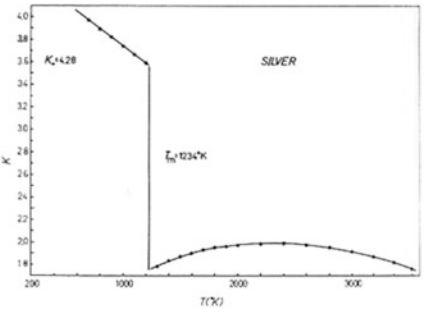

LEAD
K_s=0.355
T_m=600.6 °K
LIQUID
K
T (°K)

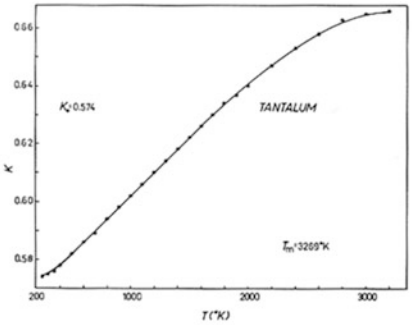

MOLYBDENUM
K_s=1.39
T_m=2883 °K
K
T(°K)

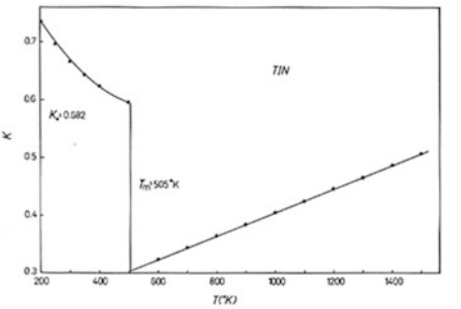

NICKEL
K_s=0.94
T_m=1720 °K
K
T(°K)

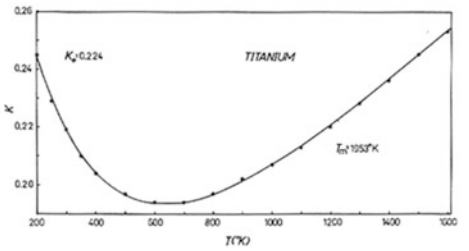

PLATINUM
K_s=0.715
T_m=2042 °K
K
T(°K)

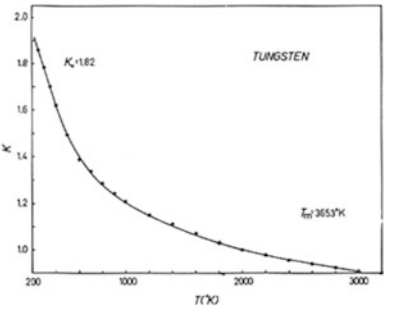

RHODIUM
K_s=1.51
T_m=2233 °K
K
T(°K)

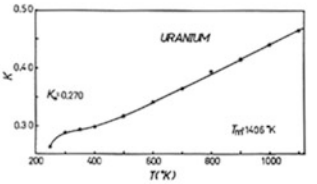

SILVER
K_s=4.28
T_m=1234 °K
K
T(°K)

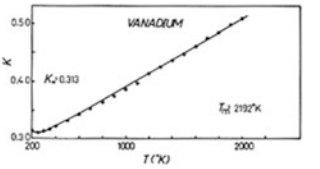

TANTALUM
K_r=0.574
T_m=3269°K
K
T(°K)

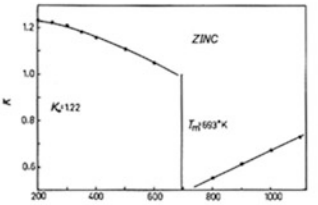

TIN
K_r=0.582
T_m=505°K
K
T(°K)

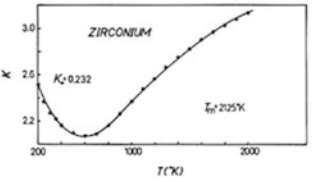

TITANIUM
K_r=0.224
T_m=1053°K
K
T(°K)

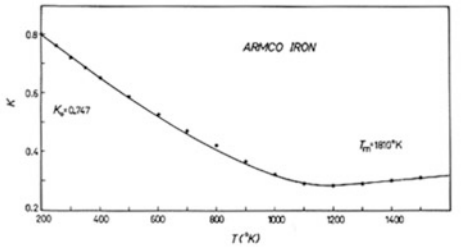

TUNGSTEN
K_r=1.82
T_m=3653°K
K
T(°K)

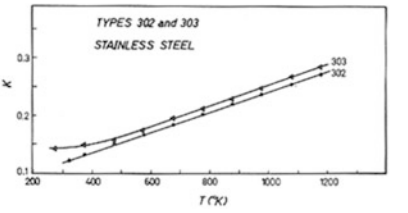

URANIUM
K_r=0.270
T_m=1406°K
K
T(°K)

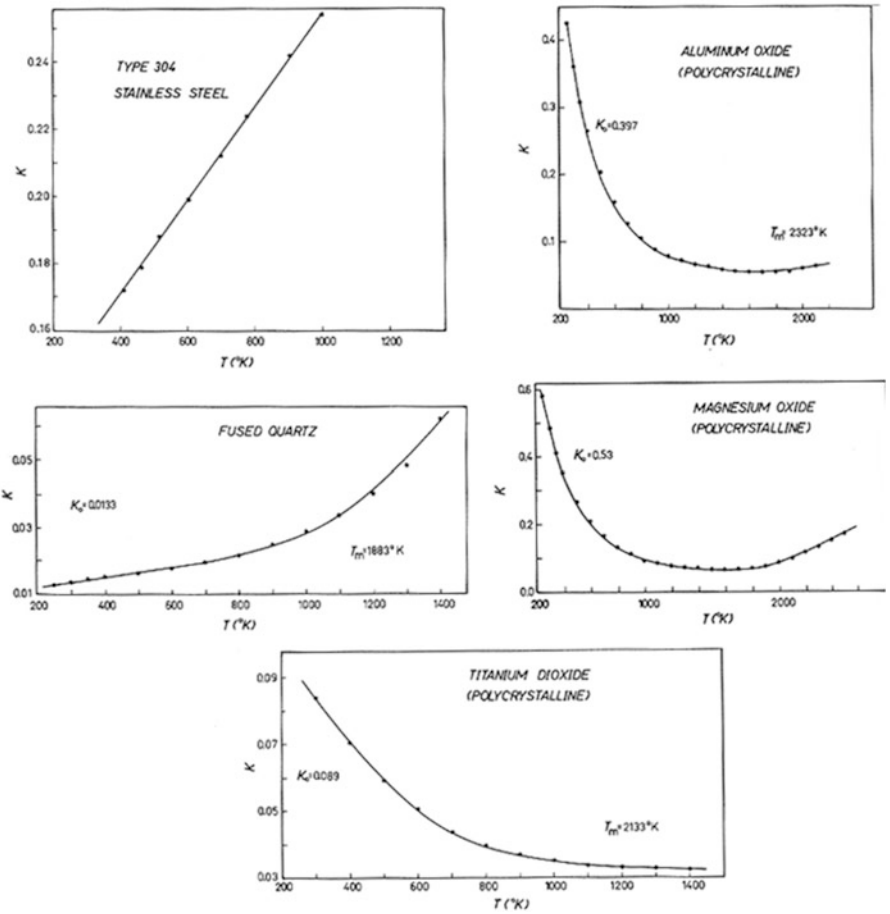

VANADIUM
K_r=0.310
T_m=2192°K
K
T(°K)


ZINC
K_r=1.22
T_m=693°K
K


ZIRCONIUM
K_r=0.232
T_m=2125°K
K
T(°K)

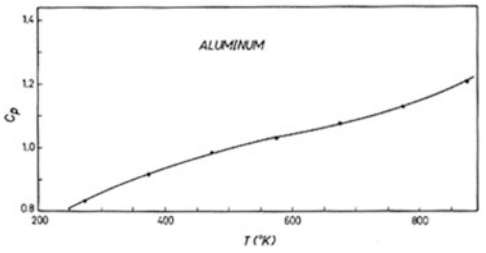

ARMCO IRON
K_r=0.747
T_m=1810°K
K
T(°K)

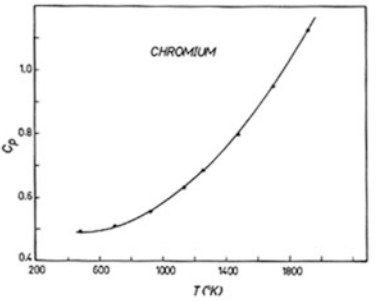

TYPES 302 and 303
STAINLESS STEEL
303
302
K
T(°K)

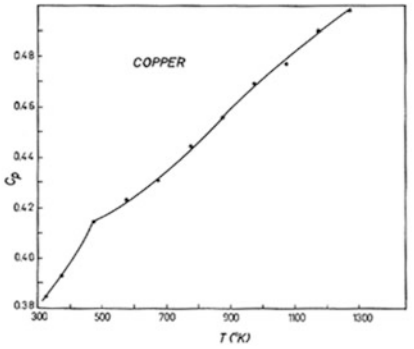


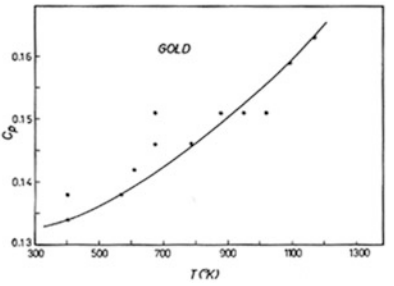
I.2 Heat Capacities

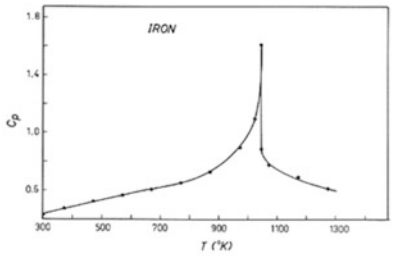
Plots of C_p versus T are given for aluminum, chromium, copper, gold, iron, lead, molybdenum, nickel, platinum, silver, tantalum, tin, titanium, tungsten, uranium, vanadium, zinc, zirconium, Armco iron, 304 stainless steel, aluminum oxide, magnesium oxide, fused quartz, and titanium dioxide. C_p is in units of J/g °C.

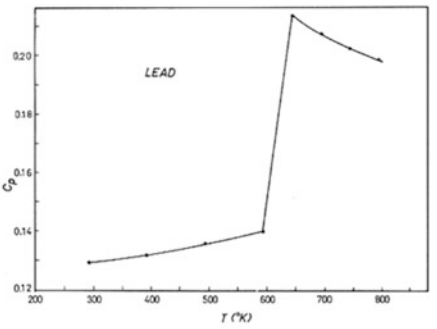

ALUMINUM
Cp
T (°K)

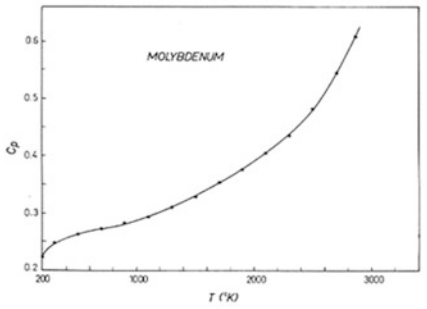

CHROMIUM
Cp
T (°K)

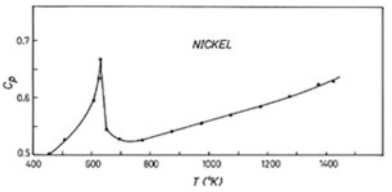

COPPER
Cp
T (°K)

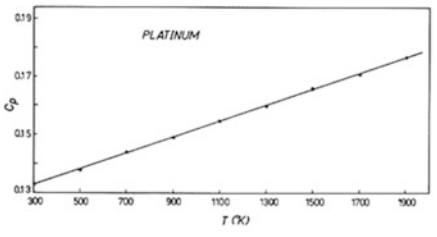

GOLD
Cp
T (°K)

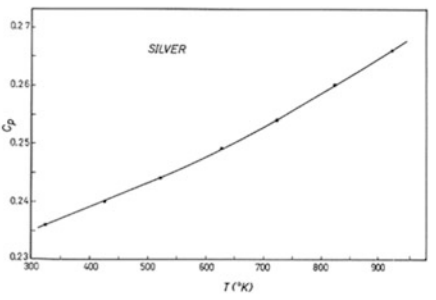

IRON
Cp
T (°K)

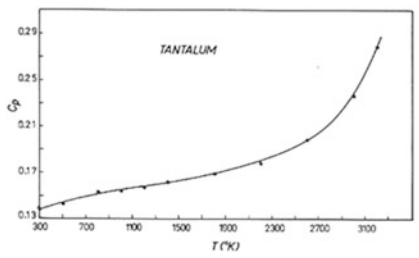

LEAD
Cp
T (°K)

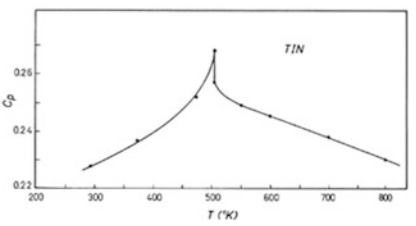

MOLYBDENUM
Cp
T (°K)

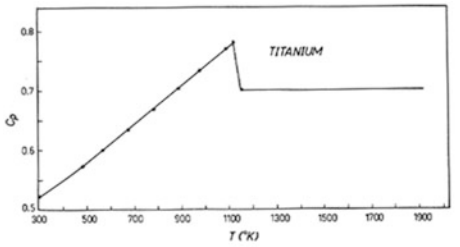

NICKEL
Cp
T (°K)

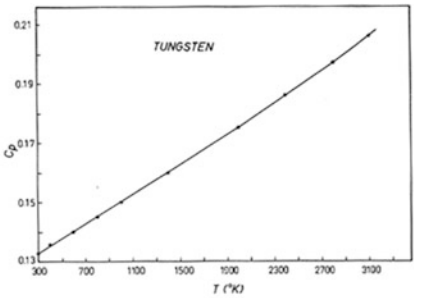

PLATINUM
Cp
T (°K)

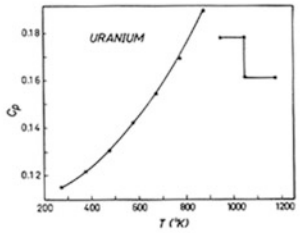

SILVER
Cp
T (°K)

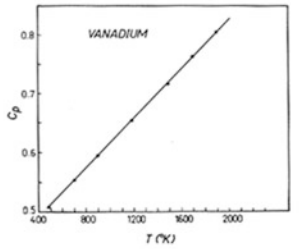

TANTALUM
Cp
T (°K)

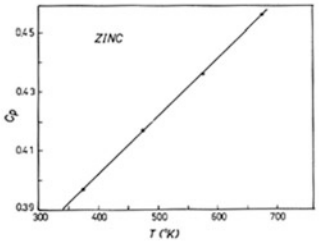

TIN
Cp
T (°K)

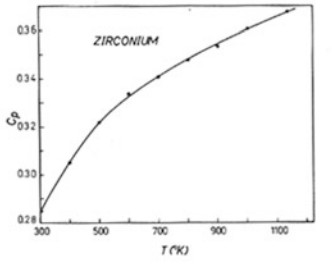

TITANIUM
Cp
T (°K)

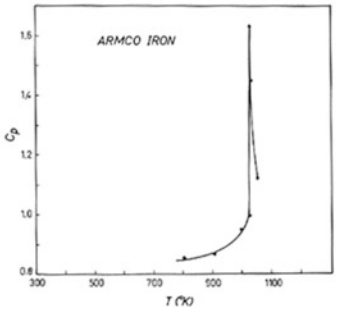

TUNGSTEN
Cp
T (°K)

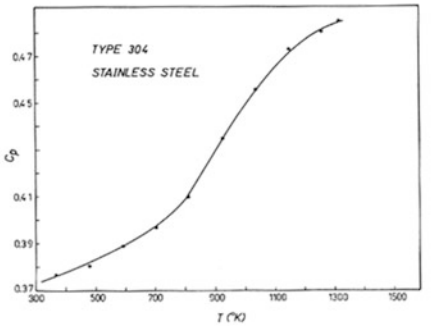

URANIUM
Cp
T (°K)

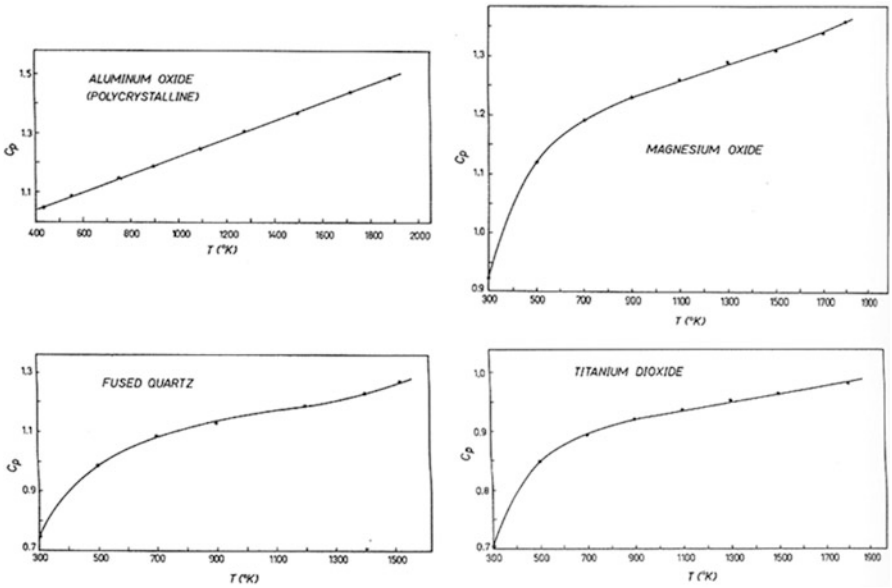

VANADIUM
Cp
T (°K)


ZINC
Cp
T (°K)


ZIRCONIUM
Cp
T (°K)

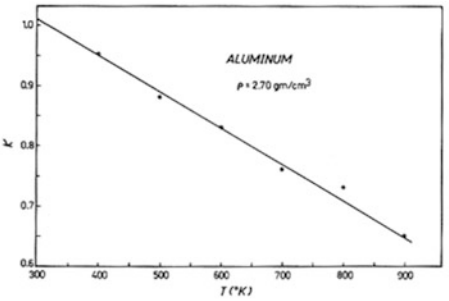

ARMCO IRON
Cp
T (°K)

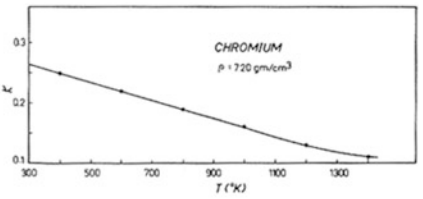

TYPE 304
STAINLESS STEEL
Cp
T (°K)

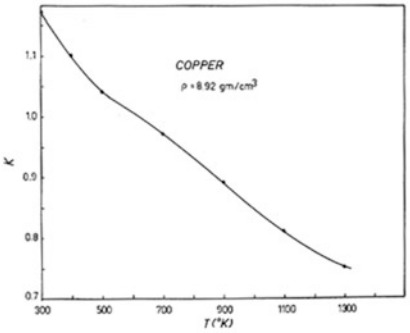


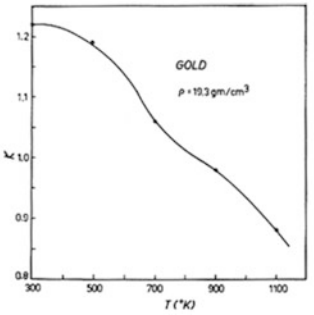
I.3 Thermal Diffusivity Data

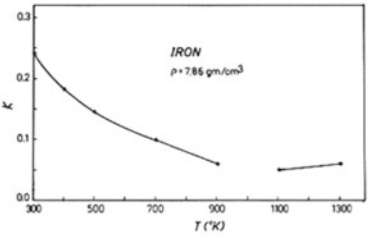
Plots of κ versus T are given for aluminum, chromium, copper, gold, iron, lead, molybdenum, nickel, platinum, silver, tantalum, titanium, tungsten, uranium, vanadium, zinc, zirconium, 304 stainless steel, aluminum oxide, magnesium oxide, fused quartz, and titanium dioxide. κ in units of cm^2/s.

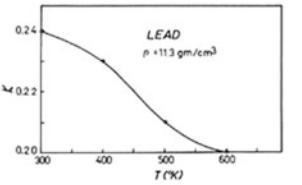

ALUMINUM
ρ = 2.70 gm/cm³
K
T (°K)

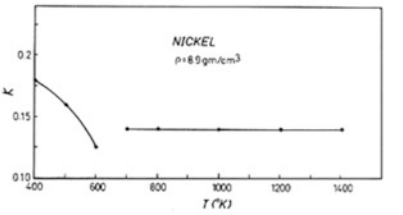

CHROMIUM
ρ = 7.20 gm/cm³
K
T (°K)

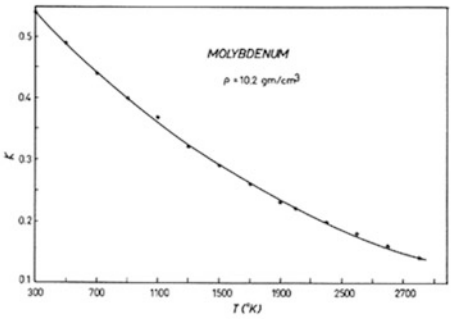

COPPER
ρ = 8.92 gm/cm³
K
T (°K)

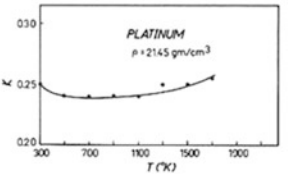

GOLD
ρ = 19.3 gm/cm³
K
T (°K)

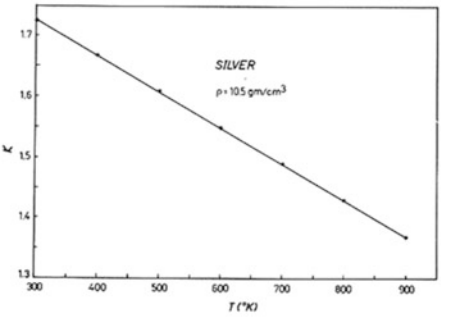

IRON
ρ = 7.86 gm/cm3
K
T (°K)

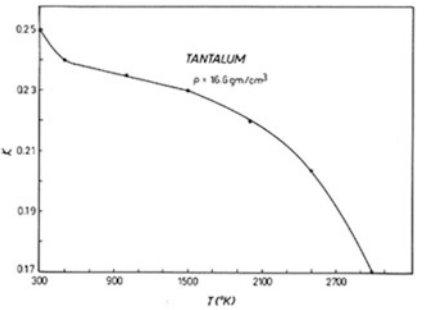

LEAD
ρ = 11.3 gm/cm3
K
T (°K)

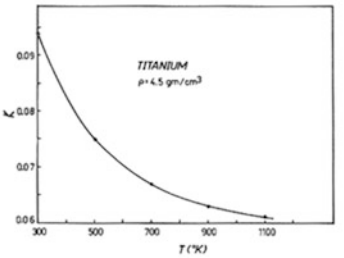

NICKEL
ρ = 8.9 gm/cm3
K
T (°K)

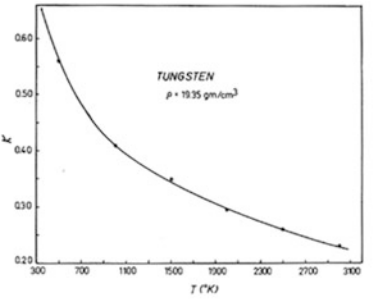

MOLYBDENUM
ρ = 10.2 gm/cm3
K
T (°K)

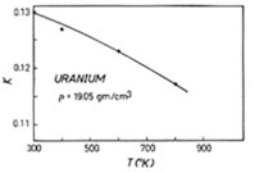

PLATINUM
ρ = 21.45 gm/cm3
K
T (°K)

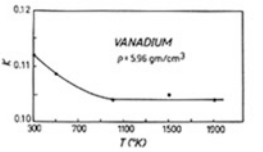

SILVER
ρ = 10.5 gm/cm3
K
T (°K)

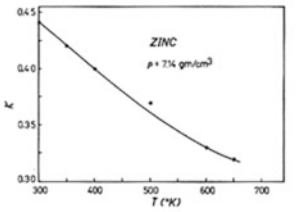

TANTALUM
ρ = 16.6 gm/cm3
K
T (°K)

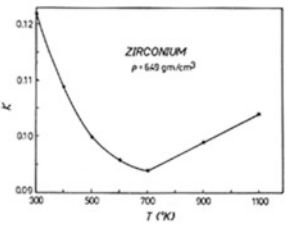

TITANIUM
ρ = 4.5 gm/cm³
K
T (°K)

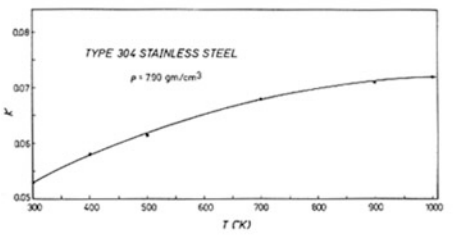

TUNGSTEN
ρ = 19.35 gm/cm³
K
T (°K)

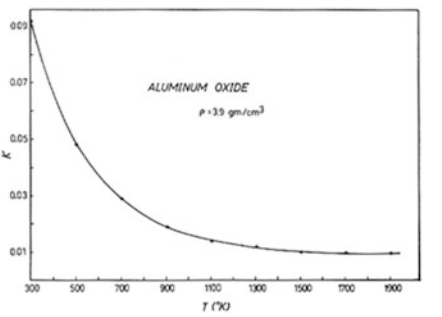

URANIUM
ρ = 19.05 gm/cm³
K
T (°K)

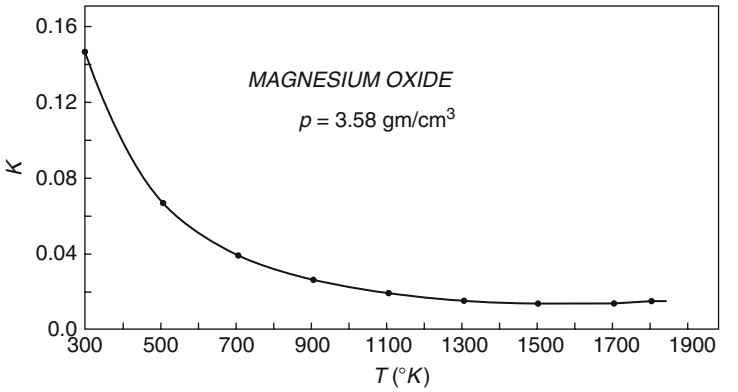

VANADIUM
ρ = 5.96 gm/cm³
K
T (°K)

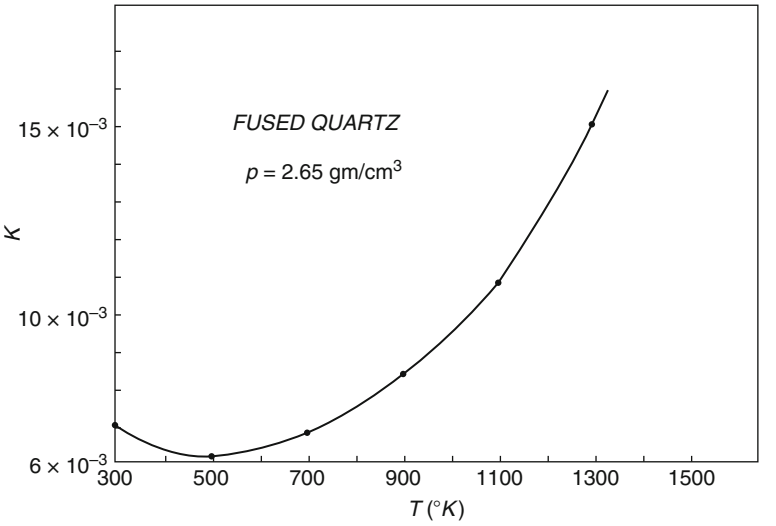

ZINC
ρ = 7.14 gm/cm³
K
T (°K)

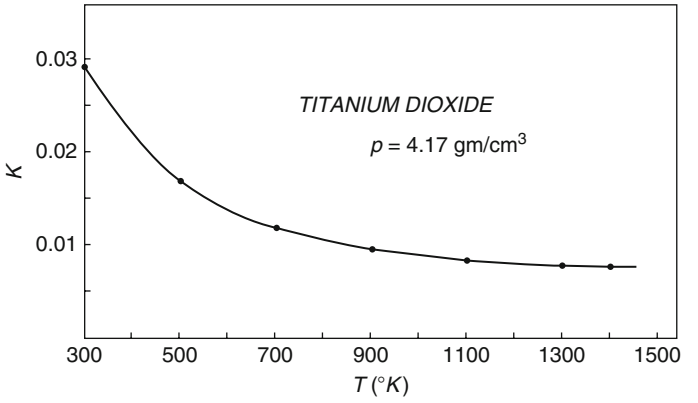

ZIRCONIUM
ρ = 6.49 gm/cm³
K
T (°K)


TYPE 304 STAINLESS STEEL
ρ = 7.90 gm/cm³
K
T (°K)


ALUMINUM OXIDE
ρ = 3.9 gm/cm³
K
T (°K)


MAGNESIUM OXIDE
p = 3.58 gm/cm3
K
0.16
0.12
0.08
0.04
0.0
300
500
700
900
1100
1300
1500
1700
1900
T (°K)


FUSED QUARTZ
p = 2.65 gm/cm3
K
15 × 10−3
10 × 10−3
6 × 10−3
300
500
700
900
1100
1300
1500
T (°K)


TITANIUM DIOXIDE
p = 4.17 gm/cm3
K
0.03
0.02
0.01
300
500
700
900
1100
1300
1500
T (°K)

I.4 References

Additional references for more information for the readers that are interested to study further

Anisimov SI et al. (1971) Effects of High-power Radiation on Metals
Knudsen M (1909) Ann. Phys., Lpz. 28 999
Finke BR, Simon G (1988) On the gas kinetics of laser-induced evaporation of metals. J Phys D: Appl Phys 23 (1990) 67–74
Knight CJ (1979) Theoretical modeling of rapid surface vaporization with back pressure. AIAA J 17(5)
Wenwu Zhang Y. Lawrence Yao (2000) Modeling and analysis of UV laser micro-machining of copper. ICALEO 2000
Crank J (1984) Free and moving boundary problems. Clarendon Press: Oxford
Ramachandran N, Gupta JR Jalunu Y (1982) Theraml and fluid flow effects during solidification in a rectangular cavity. Int J Heat Mass Transfer 25: 187–194
Gadgil A, Gobin D (1984) Analysis of two dimensional melting in rectanguar enclosures in presence of convection. J Heat Transfer 106: 20–26
Partankar SV (1980) Numerical heat transfer and fluid flow. Hemisphere: Washington
Vollwe VR, Praksh C (1987) A fixed grid numerical modeling methodology for convection-diffusion mushy region phase-change problems. Int J Heat Mass Transfer 30(8) 1709–1719
Wenwu Zhang Y, Lawrence Yao and Kai Chen (2000) Modeling and analysis of UV laser micro-machining of copper. ICALEO 2000
Patankar SV (1980) Numerical heat transfer and fluid flow. McGraw: New York
Mazumder J, Mohanty PS, Kar A (1996) Mathematical modelling of laser materials processing. Int J Mater Product Technol 11: 193–252
Frank PI, David PD (1996) Fundamentals of heat and mass transfer. John Wiley & Sons, Inc., 4th edition, New York
Frank MW (1999) Fluid Mechanics. WCB/McGraw-Hill, 4th edition, New York
Aden M et al. (1992) Laser-induced vaporization of a metal surface. J Phys D 25, 57–65
Dabby FW, Paek UC (1972) High-intensity laser-induced vaporization and explosion of solid material. IEEE J Quantum Electronics QE-8(2) 106–111
Ho JR et al. (1995) Computational model for the heat transfer and gas dynamics in the pulsed laser evaporation of metals. J Appl Phys 78(7) 4696–4709
Mazumder KA (1994) Mathematical model for laser ablation to generate nanoscale and submicometer-size particles. J. Phys Rev E, 49(1), 410–419.
Kezhun Li and Paul Sheng (1995) Computational model for laser cutting of steel plates. MED-Vol.2-1/MH-MH-Vol.3-1, Manufacturing Science and Engineering, ASME 1995, 3–14
Cai L, Sheng P (1996) Analysis of laser evaporative and fusion cutting. J Manufacturing Sci Eng, 118, 225–234
Modest MF (1996) Three-dimensional, transient model for laser machining of ablating de-composing materials. Int J Heat Mass Transfer 39(2) 221–234
Paek UC, Gagliano FP (1972) Thermal analysis of laser drilling processes. IEEE J Quantum Electronics QE-8 112–119
Singh RK, Narayan J (1990) Pulsed-laser evaporation technique for deposition of thin films: physics and theoretical model. Phys Rev B 41(3) 8843–8859

Appendix J
Acronyms and Definitions

DOF The depth of focus is the distance over which the focused beam has about the same intensity; it is defined as the distance over which the focal spot size changes −5 %~5 %.

Electronic assembly A number of electronic components (i.e., "circuit elements", "discrete components", integrated circuits, etc.) connected together to perform (a) specific function(s), replaceable as an entity and normally capable of being disassembled.

Evaporative Laser Cutting Evaporative laser cutting is the laser cutting process that target material is ablated through direct vaporization, typical applications are laser cutting of low vaporization temperature and low thermal conduction materials.

Excimer Lasers Lasers which use the noble gas compounds for lasing. Excimer lasers generate laser light in ultraviolet to near-ultraviolet spectra, from 0.193 to 0.351 μm. Gas Laser, a laser in which the active medium is a gas. The gas can be composed of molecules (like CO_2), Atoms (like He-Ne), or ions (like Ar+).

Laser Fusion cutting Laser fusion cutting is laser cutting through melting and gas jet blowing.

Ground State Lowest energy level of an atom or molecule.

Heat Affected Zone Heat affected zone is the region close to the laser-irradiated area where obvious temperature change from original area happens, or obvious strain state change happens.

Hologram An interference phenomenona captured on a plate (or film). It can contain enormous amount of information and a three-dimensional image can be constructed from it.

Knudesen layer In laser processing, strong evaporation occurs. The gas near the phase interface is not in translational equilibrium and the translational equilibrium is achieved within a few mean free paths by collisions between particles in a thin region. This region is called Knudsen layer

Laser Laser is the acronym of Light Amplification by Stimulated Emission of Radiation. Laser is light of special properties, and light is electromagnetic

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

(EM) wave in visible range. Lasers, broadly speaking, are devices that generate or amplify light, just as transistors generate and amplify electronic signals at audio, radio, or microwave frequencies. Here light must be understood broadly, since lasers have covered radiation at wavelengths ranging from infrared range to ultraviolet and even soft X-ray range.

Laser machining Laser machining is material removal accomplished by laser–material interaction, generally speaking, these processes include laser drilling, laser cutting and laser grooving, marking or scribing.

Laser Mode Laser mode is the possible standing EM waves in laser cavity.

Longitudinal (Axial) Modes Axial standing EM waves within the laser cavity.

Laser Resonator or Laser Cavity The optical mirrors, active medium and pumping system form the laser resonator, which is also called Laser Cavity. Laser cavities can be divided into Stable Cavities and Unstable Cavities according to whether they make the oscillating beam converge into the cavity or spread out from the cavity.

Line width The line width of laser is the width of laser beam frequency. Laser line width is much narrower than normal light.

Liquid Laser Lasers which use large organic dye molecules as the active lasing medium.

M^2 of the beam M^2 is a beam quality index that measures the difference between an actual beam and the Gaussian beam.

Matrix A substantially continuous phase that fills the space between particles, whiskers or fibers.

Marangoni Mechanism Liquid surface force due to temperature gradient (thermal) or composition gradient (chemical)

Microcircuit A "monolithic integrated circuit" or "multichip integrated circuit" containing an arithmetic logic unit (ALU) capable of executing a series of general purpose instructions from an external storage. N.B.1: The "microprocessor microcircuit" normally does not contain integral user-accessible storage, although storage present on-the-chip may be used in performing its logic function. N.B.2: This definition includes chip sets which are designed to operate together to provide the function of a "microprocessor microcircuit".

Multichip A "integrated circuit" where two or more "monolithic integrated circuits" bonded to a common "substrate".

Mode Locking A method to create very short laser pulses. It makes the phase difference of many modes (frequencies) in the laser cavity fixed, or locked, and thus very narrow pulses (in time) are created.

Mushy region Phase changes happen over a temperature region in general, and thus solid and liquid states coexists during phase changes. The region of this mixture of solid and liquid is called Mushy region.

Photon The minimum quantity of light energy that can be exchanged is called a light quantum or photon.

Polarized Light If the light has a dominant direction of the E vector, we say the light is polarized. Natural light is not polarized, while laser beam is polarized. Polarization can be created and adjusted by a polarizer.

Population Inversion Normally the number of atoms at high-energy level(E1) is less than those in low-energy level(E1), N2(E2) < N1(E1). If N2>N1, we say population inversion exists, which is a necessary condition for lasing.

Pumping The process to raise atoms from lower level to upper level is called pumping.

Q-Switching A method to create laser pulses. It modualates the Q (Quality) of laser cavity to build population inversion first, then release the accumulated energy suddenly, in this way high-energy pulses can be created.

Recombination Radiation In semiconductors, when the electrons combine with the holes, photons are emitted, and this is called Recombination Radiation. **Semiconductor Lasers** are based on this mechanism.

Resolution The least increment of a measuring device; on digital instruments, the least significant bit. (Reference: ANSI B-89.1.12)

Solid State Laser A laser in which the active medium is in solid state (usually not including semiconductor lasers).

Semiconductor Lasers Lasers which use semiconductor as active medium. The majority of semiconductor materials are based on a combination of elements in the third group of the Periodic Table (such as Al, Ga, In) and the fifth group (such as N, P, As, Sb), hence referred to as the III–V compounds.

Spontaneous Radiation According to quantum mechanics, the electrons of atoms can take different energy states, say E1, E2, and E3, etc., E1<E2<E3<.... Lower energy level is more stable than higher energy levels, so electrons at high-energy levels tend to decay to low-energy levels, and the energy difference between the two levels can be given out as electromagnetic radiation. This process is called Spontaneous Radiation.

Stable Cavity and Unstable Cavity Cavities can be identified as stable or unstable according to whether they make the oscillating beam converge into the cavity or spread out of the cavity, if converge it is stable, if spread out, it is unstable.

Stimulated Absorption When the atoms at lower energy levels absorb the incident energy with corresponding frequency, they jump to upper level states, and this is called Stimulated Absorption.

Stimulated Emission Under the action of the incident electromagnetic field with the corresponding frequency, the atoms at upper level have a certain possibility to jump to the corresponding lower levels, emitting electromagnetic waves or photons with the same frequency, direction, and phase with the incident waves. This process is called Stimulated Emission.

Substrate A sheet of base material with or without an interconnection pattern and on which or within which "discrete components" or integrated circuits or both can be located.

Super alloy Nickel-, cobalt-, or iron-base alloys having strengths superior to any alloys in the AISI 300 series at temperatures over 922 K (649 °C) under severe environmental and operating conditions.

TEM Mode Transverse Electromagnetic Mode (TEM) of laser beam is called TEM mode. Three indexes are used to indicate the TEM modes. TEMplq, p is the number of radial zero fields, l is the number of angular zero fields, q is the number of longitudinal fields.

YAG Yttrium/aluminum garnet

Ultra short Pulsed Laser Laser whose pulse duration time is very short, below 1 ns, usually in the fs scale.

Index

A
Absorption coefficients, 125
Absorptivity, 91
Acoustic gradient index, 222
Adaptive optical system (AOS), 173, 175, 383
Adaptive optics, 380, 383, 394, 400, 403
Airborne laser (ABL), 2, 28, 46, 57, 68, 175, 261, 411
Airborne laser laboratory (ALL), 150
Air combat command, 59
Air force research laboratory (AFRL), 177, 384
Air force weapons laboratory (AFWL), 5
American National Standards Institute (ANSI), 45
Anomalous behavior, 149
Anti-satellite (ASAT) weapons, 148
Argon laser beam, 44
Atmospheric decision aid (ADA), 384, 409, 410
Atmospheric transmission, 51
Attenuation coefficient, 386
Avogadro's number, 262, 316

B
Back door coupling, 152
Beam waist, 164
Beam wander/jitter, 155
Beam weapons, 145
Beer-Lambert's law, 255
Bidirectional reflectivity distribution function (BRDF), 36, 44, 45
Boltzmann's constant, 165, 262, 366
Bouguer's Law, 386
Bragg condition, 185, 186
Bragg diffractor, 185
Brewster angle, 223

C
Center for devices and radiological health (CDRH), 39
Chirped pulse amplification (CPA), 227
Clapeyron-Clausius equation, 256
Cloud-free line-of-sight (CFLOS), 384, 410
COIL laser, 59
Collimation in a laser, 30
Continental U.S. (CONUS), 73
Continuous wave (CW), 32, 82, 174, 293, 374, 375
Controlled thermonuclear fusion, 102
Coupling coefficient, 357
Cross-section, 387

D
Death ray, 1, 47
de Broglie wavelength, 202
Debye temperature, 317
Defense support program (DSP), 72–74
Department of defense (DOD), 18
Department of energy (DOE), 18
Depth of focus (DOF), 228
Depth of melting, 141
Deuterium fluoride (DF), 38, 388
Diffraction, 30, 169
Dirac constant, 202
Directed energy weapon (DEW), 2, 28, 81, 149
Directional hemispherical reflectance (DHR), 193

© Springer International Publishing Switzerland 2016

B. Zohuri, *Directed Energy Weapons*, DOI 10.1007/978-3-319-31289-7

Divergence, 30–32, 148, 155, 164, 166
Drude-Lorentz theory, 239
Drude-Zener theory, 255, 293

E
Electromagnetic pulse weapons, 152
Electro optic (EO), 384, 410
Electrophoretic light scattering (ELS), 163
Equations of state (EOS), 21
Excimer laser, 135

F
Finite difference method (FDM), 264, 352
Finite element method (FEM), 264
Flash Gordon, 1
Flux density, 102
Focal spot size, 225
Fourier conduction, 329
Fourier heating model, 330
Free electron laser (FEL), 54
Free space optical (FSO), 158
Frequency-dependent dielectric constant, 126

G
Gaussian beam propagates, 164
Gaussian laser beam, 329
Geosynchronous (GEO), 72
Ground based laser (GBL), 54, 169, 260

H
Hagen-Rubens limit, 246
Hagen-Rubens reflectivity, 246
Hazardous diffuse reflection area (HDRA), 42
Hazards, 36
High energy lasers (HELs), 1, 32, 80, 411
High-energy laser system-tactical army (HELSTAR), 411
Highly elliptical orbit (HEO), 72
High power microwave (HPM), 80, 152
High power radiation laser, 201, 379
Hydrogen fluoride (HF) laser, 53

I
Incident radiation absorbed, 117
Index of refraction, 128
Infrared radiation (IR), 201, 384
Integrated product teams (IPTs), 73

K
Kill confirmation, 151
Kinetic energy weapons (KEW), 81

L
Ladar, 68
Lambert's law, 386
Laplace transform, 294, 295
Large advanced mirror program (LAMP), 181
Large optics demonstration experiment (LODE), 181
Laser beam, 30
Laser cavity, 28, 176
Laser range safety tool (LRST), 28, 36, 45, 62
Laser supported absorption (LSA), 113–115
Laser supported detonation (LSD), 120, 359, 360, 365, 367, 371
Laser weapon, 1
Latent heat, 258
Latent heat of sublimation, 333, 334
Lawrence Livermore National Laboratory (LLNL), 14, 17, 23
Lidar, 68
Localized absorption, 215
Low earth orbit (LEO), 72

M
Magnetic resonance image (MRI), 18
Material parameters, 129
Maximum permissible exposures, 40, 42
Maxwell equation, 116
Maxwell's law, 257
Mechanical damage, 49
Mesosphere, 52
Metastable, 205
Mid-infrared advanced chemical laser (MIRACL), 63
Mie scattering, 43, 51
Mie scattering theory, 393
Mirror conjugation, 383
Mirror juddering, 383

N
Naval research laboratory (NRL), 155
Nd:YAG lasers, 55
Near-filed, 168
Newtonian law of cooling, 260
Night vision goggle operations weather software (NOWS), 69

Nominal ocular hazard distance (NOHD), 41
No phase change, 264
Nusselt number, 260

O
Ohm's law, 239
Optical depth, 386

P
Parallel polarization, 219
Partial differential equation (PDE), 196
Peclet number, 261
Perpendicular polarization, 219
Persian Gulf war, 1
Phase compensation instability, (PCI), 170
Phase front, 164
Phillips research site, 177
Photon correlation spectroscopy (PCS), 162
Physical damage, 149
Plank-Einstein relation, 202
Plank's constant, 366
Plank's distribution, 211
Plank's energy density distribution, 208
Plank's postulate, 210
Plank's quantization rule, 210
Plasma, 101
Plasma resonance, 246
Potential energy weapons (PEW), 81
Prandtl number, 261
Propagation efficiency, 160
Pulse repetition frequency (PRF), 32
Pulsed mode laser, 32

Q
Q-switched, 138
Q-switched laser, 107
Quasi-elastic light scattering (QLS), 162

R
Radomes (plastic radar domes), 50
Random molecular motions, 83
Rayleigh-Jeans rule, 209
Rayleigh range, 31, 32
Rayleigh scattering, 50, 51
Rayleigh's energy distribution, 208
Recession velocity, 316
Reconnaissance, surveillance, and targeting acquisition (RSTA), 147
Reduced Plank constant, 202
Reflectivity, 91
Reflectivity coefficient, 115
Refractive index, 127
Relay mirror experiment, 154
Reynolds number, 261
Ruby laser beam, 44

S
Saha equation, 366
Sandia Corporation, 20
Scientific & Technical, 384, 411
Semi-infinite solid, 264
Skin depth, 127, 207, 246
Space and strategic defense command (SSDC), 63
Space-based infrared system (SBIRS), 72
Space-based laser (SBL), 411
Spherical aberration, 228
Spot size, 206, 217
Standard temperature and pressure (STP), 353
Stefan boundary, 262
Stefan-Boltzmann constant, 261
Stephan-Boltzmann's law, 375
Strategic defense, 79
Stratosphere, 52
Stratospheric particle injection for climate engineering (SPICE), 154
Surface hardening, 91
Surface vaporization, 114

T
Tactical high energy laser (THEL), 56
Target acquisition, 67
Target acquisition weather software (TAWS), 409
Temperature-dependent reflectance of aerospace materials (TRAM), 193
Thermal blooming, 147, 168, 405
Thermal blooming effects, 150
Thermal conductivity, 105
Thermal lens, 228
Thermal parameters, 104
Thermosphere, 53
Thermophysical properties, 195
Transmissivity plus, 127
Troposphere, 52
Turbulent eddies, 156

U
Ultraviolet catastrophe, 209, 247

V
Vaporizing a target, 29
VAX/VMS computer, 21

W
Wave equation, 204
Wien's rule, 209

X
X-rays, 149